

MASTER'S THESIS

Het Leereffect van Feedback binnen de Serious Game MeTriek voor het Voorbereidend Middelbaar Beroepsonderwijs.

Van Rumund, Robin

Award date:
2021

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact us at:

pure-support@ou.nl

providing details and we will investigate your claim.

Downloaded from <https://research.ou.nl/> on date: 05. May. 2023

Open Universiteit
www.ou.nl





Het Leereffect van Feedback binnen de Serious Game MeTriek voor het Voorbereidend Middelbaar Beroepsonderwijs

The Learning Effect of Feedback within the Serious Game MeTriek for Preparatory Secondary Vocational Education

Robin van Rumund

Master Onderwijswetenschappen
Open Universiteit

Cursusnaam en cursuscode: Masterthesis – OM9906

Naam begeleider: Dr. Rob Nadolski

Datum: 12-09-2021

Inhoudsopgave

Samenvatting	3
Summary	5
1. Inleiding	7
1.1 Theoretische kader	8
1.1.1 Feedback	8
1.1.2 Feedbackniveaus en feedbacktypes	9
1.1.3 Factoren voor effectieve feedback	10
1.1.4 Feedbacktypes in het huidige onderzoek	11
1.1.5 Instructiestrategieën voor het aanleren van wiskundige vaardigheden	12
1.1.6 Game ontwerp volgens de richtlijnen van Annetta (2010)	13
1.2 Vraagstellingen en hypothesen	14
2. Pilot MeTriek	15
3. Methode	20
3.1 Ontwerp	20
3.2 Participanten	20
3.3 Materialen	21
3.4 Meetinstrumenten	22
3.5 Procedure	23
3.6 Data-analyse	24
4. Resultaten	25
4.1 Beschrijvende statistieken	25
4.2 Leerresultaten	26
4.3 Gamescores en sessieduur	29
4.4 Educatieve kwaliteit	29
5. Discussie en conclusie	30
5.1 Vergelijking met eerder onderzoek	31
5.2 Sterkte punten en beperkingen in het huidige onderzoek	31
5.3 Aanbevelingen voor vervolgonderzoek	33
5.4 Conclusie	34
Referenties	35
Bijlage	39

Het Leereffect van Feedback binnen de Serious Game MeTriek voor het Voorbereidend Middelbaar Beroepsonderwijs

Samenvatting

Serious games zijn (digitale) spellen die leren beogen te bevorderen. Serious games kunnen een positief leereffect hebben door o.a. uitdagingen, beloningen, instructie en feedback in de game te integreren (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Gee, 2003; Drijvers, 2013). Feedback wordt beschouwd als één van de vijf leerstrategieën om leerlingen te ondersteunen bij hun leerproces (William, 2011).

De meest voorkomende feedbacktypes in serious games zijn ‘Kennis van Resultaat’ (KR), ‘Kennis van Correct Resultaat’ (KCR) en ‘Elaboratieve Feedback’ (EF). Uit eerder onderzoek blijkt dat EF vaak effectiever is dan KR en KCR. Tevens blijkt dat KCR vaak effectiever is dan KR. EF is effectief bij eenvoudige en complexe denkniveaus, terwijl KR en KCR vooral effectief zijn bij eenvoudige denkniveaus, waaronder het leren van feiten.

Onderzoeken naar feedback in serious games zijn meestal uitgevoerd bij universiteiten en hogescholen, vandaar dat van der Kleij, Feskens en Eggen (2015) vraagtekens plaatsen bij de geldigheid van deze bevindingen voor het voortgezet onderwijs. Het doel van dit onderzoek is daarom nagaan of verschillende feedbacktypes (KR, KCR en EF) binnen de zelfontwikkelde serious game MeTriek resulteren in verschillende leereffecten. MeTriek betreft het leren toepassen van berekeningen met lengtematen en omtrekmaten. De verwachting was dat alle feedbacktypes een positief leereffect op de korte en de langere termijn zouden laten zien. Daarnaast werd verwacht dat EF een groter effect dan KCR zou hebben, en dat KCR een groter effect dan KR zou hebben.

Om voldoende zeker te zijn dat MeTriek afdoende educatieve kwaliteit heeft is MeTriek ontworpen op basis van het framework van Annetta (2010) dat zich richt op de borging van de educatieve kwaliteit van een serious game. De educatieve kwaliteit is gemeten met het gevalideerde instrument van Kouwenhoven (2019) in een pilot met MeTriek bij zes leerlingen op een vmbo-school met dezelfde kenmerken als de onderzoeksschool. De educatieve kwaliteit van MeTriek bleek voldoende, terwijl de feedback van de leerlingen in de pilot eveneens is meegenomen bij de concrete inrichting van het interventieonderzoek.

Het interventieonderzoek heeft een pretoets-posttoets-retentietoets ontwerp met twee experimentele groepen (EF en KCR) en een controlegroep (KR). De deelnemende leerlingen zijn aselekt over de drie onderzoeksgroepen verdeeld waarbij elk onderwijsniveau (basis, kader, gemengd) evenredig was vertegenwoordigd (N = 105). Alle leerlingen kregen een pretoets voor de voorkennis-meting. Na de pretoets speelden de leerlingen de feedbacktype-specifieke variant van MeTriek. De posttoets werd meteen na het spelen van MeTriek afgenomen, gevolgd door de licht aangepaste vragenlijst voor educatieve kwaliteit. Twee weken later werd de retentietoets afgenomen. Op basis van dit onderzoek kan worden geconcludeerd dat MeTriek voor het voorbereidend middelbaar

beroepsonderwijs (vmbo) de potentie heeft om leerlingen te laten leren, maar dat geen conclusie getrokken kan worden ten gunste van één feedbacktype. Vervolgonderzoek met verbeterde meetinstrumenten zal een antwoord moeten vinden op de vraag of elaboratieve feedback in wiskunde (serious) games zorgt voor een hoger leereffect bij leerlingen in het vmbo.

Kernwoorden: feedbacktypes, serious games, vmbo, leereffecten

Summary

Serious games are (digital) games intended to promote learning. Serious games can have a positive learning effect by integrating challenges, rewards, instruction, and feedback into the game (Connolly, Boyle, MacArthur, Hainey, & Boyle, 2012; Gee, 2003; Drijvers, 2013). Feedback is considered one of the five learning strategies to support students in their learning process (Wiliam, 2011).

The most common feedback types in serious games are 'Knowledge of Result' (KR), 'Knowledge of Correct Result' (KCR) and 'Elaborative Feedback' (EF). Previous research has shown that EF is often more effective than KR and KCR. Research has also shown that KCR is often more effective than KR. EF is effective for both simple and complex thinking levels, while KR and KCR are especially effective for simple thinking levels, including learning facts.

Studies on feedback in serious games have mostly been conducted at universities and colleges, hence van der Kleij, Feskens and Eggen (2015) question the validity of these findings for secondary education. Therefore, the aim of this study was to investigate whether different feedback types (KR, KCR and EF) within the self-developed serious game MeTriek result in different learning effects. MeTriek is about learning to apply calculations with measures of length and perimeter. All feedback types were expected to show a positive learning effect in the short and longer term. In addition, EF was expected to have a larger effect than KCR, and KCR was expected to have a larger effect than KR.

In order to be sufficiently certain that MeTriek has sufficient educational quality, MeTriek was designed based on the framework of Annetta (2010), which focuses on assuring the educational quality of a serious game. The educational quality was measured with the validated instrument of Kouwenhoven (2019) in a pilot with MeTriek with six students at a pre-vocational-school with the same characteristics as the research school. The educational quality of MeTriek proved to be sufficient, and the input from the pupils in the pilot was also taken into account in the actual design of the intervention research.

The intervention study has a pretest-posttest retention design with two experimental groups (EF and KCR) and a control group (KR). The participating students were randomly assigned to the three research groups with each educational level (primary, secondary, mixed) equally represented (N = 105). All pupils completed a pretest for the knowledge measurement. After the pretest, the pupils played the feedback type-specific variant of MeTriek. The posttest was administered immediately after playing MeTriek, followed by the slightly modified educational quality questionnaire. Two weeks later, all participants completed the retention test. Based on this study, can be concluded that MeTriek for pre-vocational secondary education has the potential to make students learn, but that no conclusion can be drawn in favor of one feedback type. Follow-up research with improved measurements might find an answer whether elaborative feedback in mathematical (serious) games increases the learning results of pre-vocational students.

Keywords: feedback types, serious games, pre-vocational, learning effects

1. Inleiding

Het bevorderen van wiskundevaardigheden is belangrijk vanuit onderwijskundige en maatschappelijke perspectieven, omdat slechte rekenvaardigheden het moeilijk maken om effectief te functioneren in de maatschappij (Parsons & Bynner, 2006). Juist omdat wiskundevaardigheden belangrijk zijn is het zorgwekkend dat het aandeel 15-jarigen met weinig wiskundevaardigheden in Nederland de laatste jaren is gestegen van 10,9 % in 2003 tot 16,7% in 2018 (Ministerie van Onderwijs, Cultuur en Wetenschap, 2019). In het reken- en wiskunde onderwijs worden steeds vaker games, zoals ‘Startrekenen 3D’ van uitgeverij Deviant gebruikt als medium voor leren. Dergelijke games verschillen van entertainment games omdat ze beogen de beheersing van wiskunde te bevorderen. Games met de intentie om de speler te laten leren, worden serious games genoemd (Johnson, Bailey, & Van Buskirk, 2017). Onderzoek toont aan dat serious games voor het vak wiskunde een positief effect kunnen hebben op leerresultaten (Wilson, Dehaene, Dubois, & Fayol, 2009; Kiili, Moeller, & Ninaus, 2018; Gaggi & Petenazzi, 2019).

Hoewel docenten en onderzoekers hoge verwachtingen hebben van het positieve effect van serious games op leerresultaten en motivatie, laten recente onderzoeken ook negatieve effecten of geen effecten zien op leerresultaten en motivatie (Clark, Tanner-Smith, & Killingsworth, 2016; Kebritchi, Hirumi, & Bai, 2010; Wouters, Nimwegen, Oostendorp, & Spek, 2013). Deze tegenstrijdige bevindingen komen vooral door de verschillen in educatieve kwaliteit tussen de serious games. Zo verschillen serious games bijvoorbeeld in gebruikte feedback, gehanteerde instructie, de relevantie van het verhaal voor de speler, de beoogde leerdoelen en speelduur.

Onderzoek heeft aangetoond dat serious games een positief leereffect kunnen hebben door uitdagingen, feedback, beloningen, instructie, leerdoelen en een educatieve context te integreren (Connolly et al., 2012; Gee, 2003; Drijvers, 2013). Het is daarom van belang om bij het onderzoek met een serious game te borgen dat deze game afdoende educatieve kwaliteit heeft om van te kunnen leren. Deze insteek is gehanteerd bij de ontwikkeling van de serious game “MeTriek” die in dit scriptie-onderzoek is gebruikt.

Eén van de voordelen van digitale systemen, zoals o.a. een serious game, is de mogelijkheid om *just in time* feedback toe te voegen (Dalziel, 2001; Ma, Adesope, Nesbit, & Liu, 2014). Feedback wordt beschouwd als een bron van informatie om leerlingen te helpen bij het herstructureren van hun conceptuele kennis en het verbeteren van hun leerproces (Hsieh & O’Neil, 2002; Hsu & Tsai, 2013).

De meest voorkomende feedbacktypes in serious games zijn ‘Kennis van Resultaat’ (KR), ‘Kennis van Correct Resultaat’ (KCR) en ‘Elaboratieve Feedback’ (EF), waarvan KR het meest wordt gebruikt (Hsieh & O’Neil, 2002; Shute, 2008). KR feedback laat de leerling weten of een antwoord goed of fout is, maar geeft niet het juiste antwoord, terwijl KCR feedback wel het juiste antwoord geeft. EF feedback geeft altijd extra informatie, maar toont niet altijd het antwoord. Onderzoeken die met EF zijn gedaan hebben verschillende operationalisaties van EF gebruikt (Shute, 2008), hierdoor zijn de

onderzoeken onderling niet goed te vergelijken. KR geeft geen informatie over hoe een leerling zichzelf kan verbeteren en is daarom het minst effectief in vergelijking met de twee andere feedbacktypes, terwijl EF over het algemeen het meest effectief is (van der Kleij et al., 2015).

Onderzoek naar feedback in serious games richt zich vooral op correctieve feedback, waardoor andere feedbacktypes minder onderzocht zijn (van der Kleij et al., 2015). Daarnaast zijn de meeste onderzoeken naar feedback in serious games uitgevoerd bij studenten van universiteiten en hogescholen, vandaar dat van der Kleij et al. (2015) vraagtekens plaatsen bij de geldigheid van de bevindingen bij deze onderzoeken voor het voortgezet onderwijs (VO). De bevindingen bij feedbackonderzoek in serious games bij universitaire en hogeschoolstudenten zijn waarschijnlijk niet van toepassing voor leerlingen in het voortgezet middelbaar beroepsonderwijs (vmbo), omdat deze populaties sterk verschillen in onderwijsniveau. Van der Kleij et al. (2015) geeft aan dat feedback bij scholieren (basisschool en voortgezet onderwijs) anders kan werken dan bij studenten (hogeschool en universiteit) en dat op universiteiten en hogescholen hogere effectgroottes zijn gevonden dan op scholen voor voortgezet onderwijs en basisonderwijs. Serious games in het (wiskunde)onderwijs voldoen vaak niet aan de eisen voor educatieve kwaliteit, zoals het integreren van uitdagingen, feedback, beloningen, instructie en leerdoelen (Connolly et al., 2012; Gee, 2003; Kiili, Devlin, & Multisilta, 2015).

Het doel van dit onderzoek is nagaan of verschillende feedbacktypes (KR, KCR, EF) binnen MeTriek resulteren in verschillende leereffecten op het gebied van kennistoepassing. Daarvoor zijn de feedbacktype-specifieke varianten van de game MeTriek in het vmbo onderzocht. MeTriek gaat over het wiskundig onderwerp ‘lengtematen en omtrek’ waarin het toepassen van nieuw geleerde kennis bij berekeningen (lengtematen en omtrek) centraal staat. De game MeTriek is door de onderzoeker ontwikkeld aan de hand van het framework van Annetta (2010) en vakdidactische expertise. Volgens de (vernieuwde) taxonomie van Bloom valt het onthouden, begrijpen en toepassen van kennis, zoals het onderwerp in deze scriptie, onder de lagere orde vaardigheden (Bruyckere, 2017).

In dit onderzoek is EF geoperationaliseerd door na het beantwoorden van een vraag aan te geven of het antwoord goed of fout is en de juiste oplossingsstrategie te tonen. Deze vorm van EF wordt ook wel EF *response contingent* genoemd (Shute, 2008).

1.1 Theoretische kader

1.1.1 Feedback

Feedback is informatie die gegeven wordt om de kloof tussen wat een leerling beheerst en niet beheerst te overbruggen (Hattie & Timperley, 2007; Wiliam, 2011). Volgens Wiliam (2011) is feedback één van de vijf leerstrategieën om leerlingen goed te ondersteunen bij hun leerproces. Eerder feedbackonderzoek geeft inzicht in de kenmerken van goede feedback, in de contextuele factoren (bijvoorbeeld: taakkenmerken, leerlingkenmerken), en in de interacties hiertussen. Het

eerdere onderzoek laat zien dat feedback over het algemeen een positief effect heeft op leerresultaten en motivatie, maar dat verkeerd uitgewerkte of verkeerd ingezette feedback geen of zelfs een negatief effect op leerresultaten en motivatie kan hebben (Hattie & Timperley, 2007; Shute, 2008; Fyfe, Ritte-Johnson, & DeCaro, 2012; van der Kleij et al., 2015).

1.1.2 Feedbackniveaus en feedbacktypes

Hattie en Timperley (2007) onderscheiden vier niveaus van feedback en claimen dat het niveau van feedback direct invloed heeft op de effectiviteit van de feedback. Het eerste niveau is taakgerichte feedback die zich richt op de taak of het product, en is de eerder genoemde KR en KCR. Een voorbeeld van taakgerichte feedback is: “Je hebt het juiste antwoord gekozen”. Het tweede niveau is procesgerichte feedback. Procesgerichte feedback richt zich op het oplossingsproces en wordt tijdens het maken van de taak gegeven zodat een leerling deze kan afronden. Procesgerichte feedback kan dienen als aanwijzing, gerelateerd aan de taak, en kan leiden tot het gebruik van betere oplossingsstrategieën om een taak uit te voeren. Een voorbeeld van procesgerichte feedback is: “Je hebt het antwoord niet goed, je had het beter op manier X kunnen aanpakken”. Het derde niveau is feedback op het niveau van zelfregulatie. Een voorbeeld is: “De stappen die je moet zetten om tot het juiste antwoord te komen zijn (uitleg)”. Tot slot bestaat het vierde niveau van feedback uit feedback die zich richt op de leerling en niet op de taak.

Behalve over feedbackniveaus spreekt de literatuur ook van verschillende feedbacktypes. Naast de feedbacktypes KR en KCR vermeldt de literatuur: ‘*no feedback*’, ‘*try again feedback*’ en ‘Elaboratieve Feedback’ (EF) (Shute, 2008). De leerling krijgt bij *try again feedback* de kans om de vraag bij een fout antwoord nog een keer te beantwoorden.

Voor het geven van EF bestaan meerdere varianten waarin (meestal) wordt aangegeven of het antwoord juist is en wordt vaak extra informatie gegeven (Shute, 2008; van der Kleij et al., 2015). Elaboratieve feedback kan zich volgens van der Kleij et al. (2015) door de extra gegeven informatie richten op alle vier de feedbackniveaus. Shute (2008) maakt onderscheid in feedback op basis van complexiteit. De minst complexe feedbacktypes zijn KR, KCR en ‘*try again*’.

EF is volgens Shute (2008) in te delen in zes feedbacktypes, oplopend in complexiteit. Bij het eerste type, *attribute isolation*, wordt informatie gegeven over concepten van de taak of vaardigheid. Het tweede type is *topic contingent* waarbij de leerling informatie krijgt over het onderwerp dat op dat moment aan bod komt. *Response contingent*, ook wel extra-instructiefeedback genoemd, biedt verificatie en een vraagspecifieke uitwerking (Mason & Bruning, 2001).

De leerling krijgt bijvoorbeeld extra instructie in de vorm van een uitgewerkt voorbeeld over de vraag aangeboden. Als het antwoord niet expliciet wordt gegeven, maar feedback wordt gegeven om de leerling naar het juiste antwoord te sturen, dan noemt men dit *hints, cues of prompts*. Het vijfde type (*bugs en misconceptions*) gaat specifiek in op fouten of misconcepties. Dit feedbacktype vraagt veel van de docent, omdat een foutenanalyse en een diagnose nodig is. Het laatste, meest complexe

feedbacktype is *informative tutoring*. Bij dit feedbacktype wordt gebruik gemaakt van KR, *error flagging* en *hints*.

Shute (2008) geeft net zoals Mason en Bruning (2001) aan dat onder andere *timing* (direct of uitgesteld), het leerniveau van de leerling, en de complexiteit van de inhoud van de feedback aandachtspunten zijn voor het geven van feedback. Als de inhoud van de feedback te complex is voor het leerniveau van de leerlingen, dan zullen leerlingen daar volgens Shute (2008) geen aandacht aan besteden.

1.1.3 Factoren voor effectieve feedback

Onderzoek toont aan dat feedback over het algemeen een positief effect heeft op leerresultaten en motivatie (Hattie & Timperley, 2007; Shute, 2008; van der Kleij et al., 2015). Volgens Hattie en Timperley (2007) is feedback gericht op de persoon het minst effectief. Bovendien kan op de persoon gerichte feedback een negatief effect op leren hebben omdat dit kan leiden tot aangeleerde hulpeloosheid of sociale vergelijking (Hattie & Marsh, 1995). Taakgerichte feedback is effectief voor het verbeteren van oplossingsstrategieën of het verbeteren van zelfregulatie en dit feedbacktype is vooral bruikbaar bij foute interpretaties van de lesstof (Hattie & Timperley, 2007). Procesgerichte feedback en feedback op zelfregulatie is vooral effectief voor het dieper verwerken van de stof en voor de beheersing van vaardigheden. De effectiviteit van feedback op het niveau van zelfregulatie hangt af van het vermogen tot zelfassessment en zelfevaluatie, de bereidheid om gebruik te maken van feedback, het zelfvertrouwen van een leerling in het gegeven antwoord, de attributie van eigen succes of falen en de vaardigheid om hulp te zoeken. Uit de meta-analyse van van der Kleij et al. (2015) blijkt dat feedback gecombineerd op taakniveau en procesniveau over het algemeen het grootste effect heeft.

Kijkend naar de effecten van de verschillende feedbacktypes, blijkt dat EF vaker effectiever is dan KR en KCR (Shute, 2008; Butler, Godbole, & Marsh, 2013; van der Kleij et al., 2015) en dat KCR effectiever is dan KR. Het blijkt ook dat extra uitleg in combinatie met KCR een groter effect kan hebben dan extra uitleg zonder KCR. Timmers en Veldkamp (2011) vonden een positief verband tussen sessieduur en de aandacht die de leerlingen besteden aan de feedback. Studenten besteedden in hun onderzoek meer tijd aan de toelichting in de feedback bij een fout beantwoorde vraag dan aan deze uitgebreide feedback bij een goed beantwoorde vraag. Volgens Timmers en Veldkamp (2011) hebben studenten daarom liever feedback met extra uitleg (EF) dan feedback met alleen de vermelding van goed of fout. Uit het onderzoek van Maier, Wolf en Randler (2016) blijkt echter dat EF feedback een minder groot effect heeft dan KCR feedback. Uit onderzoek van Phye (1979) en Dempsey, Driscoll, & Lichtfield (1993) blijkt zelfs dat KR en KCR effectiever zijn bij het leren van feiten, dan het gebruik van uitgebreidere feedback. Uit eerder onderzoek blijkt ook dat bij computer gebaseerde instructie geen feedbacktype bestaat dat het beste werkt voor alle leerlingen en dat onder andere gekeken moet worden naar leerling- en taakkenmerken (Mason & Bruning, 2001; Narciss & Huth, 2004; Shute, 2008). Leerlingkenmerken zoals het prestatieniveau van de ontvanger, het vermogen van

learner control (zelfregulatie), het inschattingsvermogen of een antwoord goed (of fout) is, en de voorkennis van een leerling zijn factoren voor de effectiviteit van feedback (Mason & Bruning, 2001; Narciss & Huth, 2004; Shute, 2008).

Op het gebied van *timing* hebben leerlingen met een laag leerniveau meestal meer baat bij directe feedback en leerlingen met een hoog leerniveau hebben meestal meer baat bij uitgestelde feedback (Mason & Bruning, 2001; Shute, 2008). Daarnaast is ‘*try again*’ feedback voor leerlingen met een laag leerniveau, lage zelfregulatie en weinig inschattingsvermogen of een vraag goed (of fout) is, minder effectief dan feedback waarbij het goede antwoord wordt gegeven, zoals KCR en een vorm van EF (Shute, 2008). Leerlingen met een laag leerniveau en een lage zelfregulatie hebben meestal weinig voorkennis, laag zelfvertrouwen en weinig vermogen om zichzelf te monitoren tijdens het leren, waardoor het herhalen van een vraag kan zorgen voor frustratie en een hogere cognitieve belasting (Mason & Bruning, 2001). Feedback heeft vaak een sterk positief leereffect op leerlingen met minder voorkennis, maar heeft neutrale of zelfs negatieve leereffecten op leerlingen met meer voorkennis (Fyfe et al., 2012; Fyfe, 2016). Als een leerling te weinig voorkennis heeft om een taak te kunnen afronden is EF met extra instructie volgens Hattie en Timperley (2007) beter dan alleen KC of KCR.

De moeilijkheidsgraad van het leerdoel en het soort kennis dat in een taak gevraagd wordt, valt onder taakkenmerken. Directe feedback is effectief bij leerlingen met een laag leerniveau voor zowel eenvoudige als complexe leerdoelen en voor leerlingen met een hoog leerniveau in combinatie met eenvoudige leerdoelen (Mason & Bruning, 2001; Hattie & Timperley, 2007). Leerlingen met een hoog leerniveau en complexe leerdoelen hebben baat bij uitgestelde feedback. Bij procedurele en conceptuele kennis is directe feedback het meest effectief en voor het leren van diepere kennis en transfer is uitgestelde feedback effectief (Shute, 2008). KR en KCR hebben een klein effect wanneer het leerdoel van de les zich richt op eenvoudige denkniveaus, terwijl KCR ook een klein tot gemiddeld effect laat zien bij complexere denkniveaus. EF blijkt effectief te zijn voor zowel eenvoudige als complexe denkniveaus (Shute, 2008). Als EF te uitgebreid is, of als weinig motivatie bij de leerlingen is, of de leerlingen denken de feedback niet nodig te hebben, dan kan EF leiden tot lagere resultaten dan andere feedbacktypes. Dit kan zijn omdat de leerlingen de feedback dan niet lezen, of dat bij het wel lezen van EF het denkproces zelfs negatief wordt beïnvloed (Phye, 1979; Maier et al., 2016). Een ander nadeel van te uitgebreide EF kan zijn dat leerlingen ‘doorklikken’, omdat ze weten dat ze dan het juiste antwoord te zien krijgen (Bokhove & Drijvers, 2012). Uit eerder onderzoek is niet eenduidig af te leiden hoe uitgebreid EF moet zijn, en de operationalisatie van de ‘mate van uitgebreidheid’ is niet consistent beschreven, waardoor de bevindingen bij EF moeilijk te interpreteren zijn.

1.1.4 Feedbacktypes in het huidige onderzoek

In deze scriptie is onderzoek gedaan met verschillende feedbacktypes binnen de game MeTriek. Deze game richt zich op de ontwikkeling van rekenvaardigheden (lengtematen en omtrek berekenen) bij een vmbo doelgroep met verschillende niveaus: Basis Beroepsgerichte Leerweg (BBL), Kader

Beroepsgerichte Leerweg (KBL) en Gemengde Leerweg (GL). Het vmbo in het algemeen is het laagste diplomagerichte niveau in het Nederlandse middelbare onderwijs. In deze scriptie wordt verondersteld dat deze doelgroep een laag leerniveau en weinig vermogen tot zelfregulatie heeft in vergelijking tot andere leerlingen in het voortgezet onderwijs. Deze vmbo leerlingen gaan door het spelen van MeTriek voor de eerste keer op het vmbo aan de slag met het omrekenen van lengtematen en het berekenen van de omtrek, waardoor rekening gehouden moet worden met weinig voorkennis. Het omrekenen van lengtematen en het berekenen van de omtrek zijn voor deze leerlingen complexe vaardigheden waarbij veel procedurele kennis nodig is omdat ze kennis moeten toepassen.

Kijkend naar het framework van Mason en Bruning (2001) kan voor een doelgroep met een laag leerniveau en weinig voorkennis in combinatie met een complexe taak waarschijnlijk het beste KCR in combinatie met EF response contingent (afgekort tot EF) gebruikt worden. Hierbij geldt als kanttekening dat het framework van Mason en Bruning (2001) slechts beperkt is onderzocht.

In deze scriptie is gekozen om de feedbacktypes KCR met EF, als ook KCR te onderzoeken. Beide feedbacktypes laten een verbetering op leerresultaten zien bij complexe taken (Shute, 2008; van der Kleij et al., 2015). Hierbij is het interessant om te kijken naar de verschillen tussen feedbacktype KCR met een feedbacktype waar KCR met EF wordt gebruikt omdat uitgebreide feedback zoals EF een negatievere impact zou kunnen hebben op de ervaren educatieve kwaliteit van de game (bijvoorbeeld *flow*) in vergelijking met minder uitgebreide feedbacktypes zoals KR en KCR (Annetta, 2010; Shute, 2008). Daarnaast is het mogelijk dat leerlingen bij EF snel door de vragen gaan om bij de uitleg van het antwoord te komen, wat betekent dat ze minder nadenken over de vraag (Bokhove & Drijvers, 2012).

Leerlingen die deelnemen aan dit onderzoek zitten op het vmbo en hebben een laag leerniveau, en weinig voorkennis. Volgens Hattie en Timperley (2007) heeft deze doelgroep bij complexe taken profijt van directe feedback. In dit onderzoek krijgen de leerlingen daarom meteen na het beantwoorden van een vraag de feedback te zien. De controlegroep krijgt KR feedback aangeboden, dit feedbacktype kan een klein effect hebben op leerresultaten. In deze scriptie wordt niet gekozen voor *'try again'*, omdat dit een negatief effect kan hebben op leerlingen met een laag leerniveau (Mason & Bruning, 2001). Kortom, in dit onderzoek wordt EF, KCR en KR gebruikt waarbij feedback direct wordt gegeven.

1.1.5 Instructiestrategieën voor het aanleren van wiskundige vaardigheden

Naast feedback kunnen Instructiestrategieën binnen serious games ook van invloed zijn op de leerresultaten. Het is daarom belangrijk om een effectieve instructiestrategie voor wiskunde in MeTriek te gebruiken.

Dunlosky, Rawson, Marsh, Nathan en Willingham (2013) beschrijven tien instructiestrategieën, waarvan vijf effectief zijn voor het leren van kennis en/of vaardigheden (*'elaborative interrogation'*, *'self-explanation'*, *'practice testing'*, *'distributed practice'* en *'interleaved practice'*). Volgens

Wouters en van Oostendorp (2013) leren leerlingen vaardigheden in een serious game door te ervaren en te oefenen. Een effectieve manier om de relevante informatie actief te verwerken is het gebruik van self-explanation (Mayer, 2014). Self-explanation kan leerlingen in een serious game helpen om representaties van hun kennis te generen en het kan een positieve invloed hebben op de toegankelijkheid en de toepassing van kennis (ter Vrugte & de Jong, 2017). Onderzoek laat zien dat self-explanation aangemoedigd kan worden door (onvolledige) uitgewerkte voorbeelden te gebruiken (Atkinson & Renkl, 2007). Uitgewerkte voorbeelden helpen leerlingen bij het identificeren van relevante informatie tijdens het oplossen van een (wiskundig)probleem (Atkinson, Derry, Renkl & Wortham, 2000; Wouters & van Oostendorp, 2013). Hierbij ziet de leerling hoe de vaardigheden bij het oplossen van het probleem worden toegepast. Een uitgewerkt voorbeeld bevat minimaal een vraag, een stappenplan om bij het antwoord te komen en het antwoord op de vraag (Anderson, Fincham, & Douglass, 1997). Uitgewerkte voorbeelden worden vaak gebruikt bij het aanleren van (complexe) vaardigheden (van Merriënboer & Kirschner, 2017) en geven de leerlingen een handvat om de opdracht aan zichzelf uit te leggen, wat ook effectief is voor leren (Atkinson et al., 2000).

Voor het combineren van realistische problemen en uitgewerkte voorbeelden is het onvolledige uitgewerkte voorbeeld geïntroduceerd. Het blijkt dat het combineren van een realistisch probleem en een uitgewerkt voorbeeld samen voor betere leerresultaten kan zorgen (van Merriënboer & Kirschner, 2017).

Voor dit onderzoek is gekozen voor het gebruik van de instructiestrategie self-explanation in combinatie met uitgewerkte voorbeelden omdat deze instructiestrategie bruikbaar blijkt te zijn voor alle doelgroepen en voor het vak wiskunde (Dunlosky et al., 2013). Barton (2019) noemt het werken met uitgewerkte voorbeelden en uitleggen-aan-zichzelf tijdens wiskunde lessen een superkrachtig uitgewerkt voorbeeld. In dit scriptieonderzoek is niet gekozen voor 'onvolledige' uitgewerkte voorbeelden omdat dit bij deze doelgroep waarschijnlijk kan leiden tot cognitieve overbelasting (Sweller, Merriënboer & Paas, 1998). Deze uitgewerkte voorbeelden nemen in MeTriek toe in moeilijkheidsgraad per level. Het oplopen van de moeilijkheidsgraad is belangrijk voor het aanleren van vaardigheden omdat stof aangeboden in deelstappen makkelijker wordt onthouden (van Merriënboer & Kirschner, 2017).

1.1.6 Game ontwerp volgens de richtlijnen van Annetta (2010)

Digitale serious games zijn games om een gebruiker specifieke vaardigheden aan te leren via een elektronische/computeromgeving (Annetta, 2010). Annetta (2010) heeft een framework ontwikkeld dat bestaat uit een hiërarchisch model met zes elementen die samen het kader voor effectief leren binnen een serious game vormen.

De basis van de hiërarchie is het element *identity*, waarmee bedoeld wordt dat elke speler zich een uniek personage in het spel voelt. Het tweede element is *immersion* (onderdompeling), wat betekent dat een speler betrokken is bij de inhoud van het spel en zijn of haar aanwezigheid voelt in een

bepaalde context. Spelers worden dan betrokken en intrinsiek gemotiveerd om de game succesvol af te ronden. Dit gevoel van immersion en identity kan bijvoorbeeld worden gecreëerd door een verhaallijn in te bouwen en gebruik te maken van een non-playing character (NPC) (Kiili, 2005). Als een speler dit ervaart kan de speler in een staat van *flow* terecht komen. Volgens Finneran en Zhang (2005) raken spelers tijdens een *flow* in een staat van bewustzijn, waardoor ze extra gefocust zijn, zich niet bewust zijn van hun omgeving en hoge prestaties vertonen. Volgens Kiili (2005) is het geven van feedback en het duidelijk stellen van doelen essentieel voor *flow*. Een niet goed ontwikkeld spel of verkeerde opdrachten kunnen deze status van *flow* in de weg zitten. Bij het derde element *interactivity* gaat het erom dat een speler de kans krijgt om in het spel te communiceren met anderen. Dit kan met een computergestuurd gamepersonage zijn of met bijvoorbeeld peers. Bij effectieve serious games is sprake van het vierde element *increased complexity*. Spelers worden uitgedaagd door complexere opdrachten, maar raken snel verveeld door te eenvoudige opdrachten. Het vijfde element is *informed teaching* waar de feedback en de ingebouwde assessments in serious games onder worden verstaan. Door de computer is het voor onderzoekers mogelijk om data te verzamelen over bijvoorbeeld de speeltijd van het spel, de locatie waar het spel gespeeld wordt en de interactie tussen spelers en NPC. Het laatste element is *instructional*, waarmee bedoeld wordt dat in een serious game instructie aan bod moet komen.

Voor de ontwikkeling van de serious game in dit onderzoek is Seppo (seppo.io) gebruikt. Seppo is software waarmee relatief eenvoudig – zonder programmeerkennis - serious games kunnen worden gemaakt. Dit heeft tevens als voordeel dat de theoretische concepten uit het framework van Annetta (2010) door de onderzoeker (eenvoudig) waren toe te voegen. De verantwoording over de implementatie van de zes elementen van Annetta (2010) is te vinden in bijlage 3.

1.2 Vraagstellingen en hypothesen

In deze scriptie is eerst een pilot met het zelf ontwikkelde prototype van de game MeTriek geëvalueerd via een *cognitive walkthrough* en hardop denken waarbij het doel was te kijken of de game voldoende gebruiksvriendelijk is, en of de game beschikt over genoeg educatieve kwaliteit. In het prototype werd het feedbacktype EF gebruikt, zodat tijdens de evaluatie ook gekeken kon worden naar de complexiteit van de inhoud van de feedback. De verwachting was dat de onderdelen identity, immersion, interactivity, increasing complexity, informed teaching en instructional voldoende aan bod komen in de game MeTriek en dat geen technische problemen zouden zijn. Het prototype van MeTriek scoorde op alle onderdelen voldoende en is op basis van de antwoorden uit de evaluatie bijgesteld voor het interventieonderzoek (zie verder het hoofdstuk ‘Pilot Metriek’).

Aansluitend is het interventieonderzoek uitgevoerd. In dit interventieonderzoek is door middel van een gerandomiseerd pretoets-posttoets-retentietoets ontwerp met twee experimentele groepen (KCR, EF) en een controlegroep (KR) gekeken naar het effect van het feedbacktype op de leerresultaten van de leerlingen. In Figuur 1 is schematisch het verloop van de pilot en het interventieonderzoek

weergegeven. De feedback is geïntegreerd in een feedbacktype-specifieke variant van MeTriek. De respondenten zijn in drie groepen gerandomiseerd, waarbij de game-varianten alleen verschilden op het gebruikte feedbacktype.

De centrale vraag in dit onderzoek is: “Wat is het leereffect van de feedbacktypes KR, KCR en EF bij het spelen van de serious game MeTriek door leerlingen in het vmbo?”

Hierbij golden de volgende hypothesen:

- Hypothese 1: Het spelen van de serious game MeTriek leidt bij elk feedbacktype tot een positief leerresultaat.
- Hypothese 2a: Het spelen van de serious game MeTriek met feedbacktype EF leidt tot betere leerresultaten dan het spelen van MeTriek met feedbacktype KCR en KR.
- Hypothese 2b: Het spelen van de serious game MeTriek met feedbacktype KCR leidt tot betere leerresultaten dan het spelen van MeTriek met feedbacktype KR.

Pilot MeTriek								
Ontwerp game	Ontwerp game	Ontwerp game	Ontwerp game Werving Pilot	Werving respondenten	Pilot met prototype MeTriek	Bijstellen MeTriek	Bijstellen MeTriek	Bijstellen MeTriek
Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9

Interventieonderzoek									
Werving respondenten	Werving respondenten	Werving respondenten	Werving respondenten	Werving respondenten	Introductie MeTriek en onderzoek	Pretoets, interventie, posttoets, vragenlijst Educatieve kwaliteit	Nakijken toetsen	Nakijken toetsen	Retentietoets
Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14

Figuur 1. Fasering bij de pilot met MeTriek en het interventieonderzoek met feedbacktype-specifieke varianten van MeTriek.

2. Pilot MeTriek

Tijdens de pilot is het prototype van de eigen gemaakte game MeTriek geëvalueerd via een cognitive walkthrough en hardop denken. Het doel van deze pilot was te kijken of MeTriek voldoende gebruiksvriendelijk is, beschikt over genoeg educatieve kwaliteit, en de speelduur te bepalen.

Vanwege praktische redenen en de Covid-19 pandemie was de pilot op een andere vmbo-school dan de school voor het interventieonderzoek. In de pilot participeerden leerlingen (N = 6) met vergelijkbare kenmerken als de leerlingen van de vmbo school ‘X’ waarop het interventieonderzoek zou plaatsvinden. Daardoor bleven alle leerlingen van vmbo school ‘X’ beschikbaar voor deelname aan het interventieonderzoek.

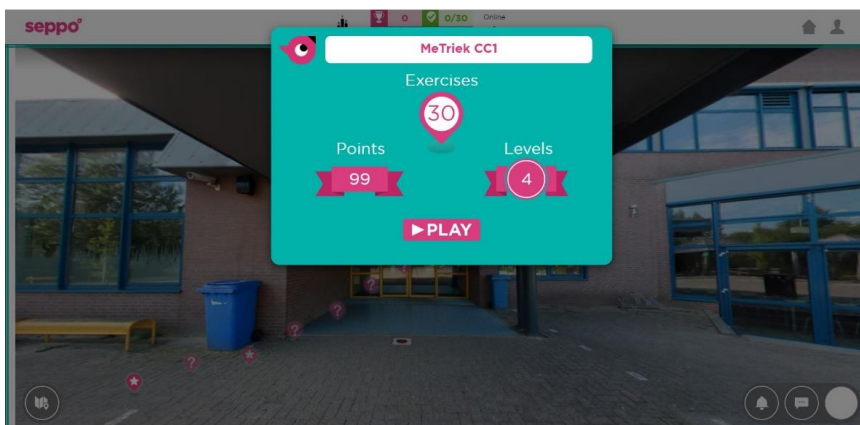
De gebruiksvriendelijkheid van MeTriek werd bepaald via een cognitive walkthrough en een nagesprek. De educatieve kwaliteit van MeTriek werd gemeten met de vragenlijst van Kouwenhoven (2019). De speelduur werd bepaald door te observeren of leerlingen bezig waren met de game en de speelduur per speler werd door het programma Seppo bijgehouden.

MeTriek is een online wiskundegame die draait op het platform Seppo. Het wiskunde-onderwerp betreft het leren toepassen van berekeningen met eenheden, lengtematen, omtrekmaten, oppervlaktematen. MeTriek is door de onderzoeker op basis van de wiskundemethode Getal en Ruimte en in samenspraak met wiskundeleraars ontwikkeld waarbij rekening is gehouden met de zes elementen van Annetta (2010). Binnen MeTriek is een verhaallijn gebruikt waarin de avatar met de naam “WisKid” de spelers van de game vraagt om te helpen. WisKid stuurt de spelers aan met tekstberichten. WisKid ziet er in de game bij elke opdracht anders uit. De veronderstelling van de onderzoeker was dat het spel daardoor interessanter is (Figuur 2).



Figuur 2. Verschillende weergave van WisKid.

MeTriek bestaat uit 30 opdrachten, verdeeld over vier levels waarvoor 99 punten gehaald kunnen worden (Figuur 3). Bij de eerste opdracht worden de in-game knoppen uitgelegd. MeTriek is gemaakt om in 10 tot maximaal 20-25 minuten uit te spelen. De levels lopen op in moeilijkheidsgraad, maar de opdrachten binnen één level zijn even moeilijk.



Figuur 3. Beginscherm met informatie.

De levels worden op voor de leerlingen bekende verschillende virtuele locaties gespeeld (schoolingang, aula, gang, moestuin) om het gevoel van *immersion* te versterken (Figuur 4).



Figuur 4. De moestuin op school.

In MeTriek zitten naast opdrachten ook theorieblokken. Theorieblokken zijn tekstblokken met uitleg gekoppeld aan een set van opdrachten binnen een level. Theorieblokken worden uniform aangegeven met dezelfde afbeelding van WisKid (Figuur 5). De onderzoeker koos bewust voor een variatie aan opdrachten om het spel interessant te houden. Zo waren er o.a. sleepvragen (Figuur 6) en meerkeuzevragen (Figuur 7).

Eenheden en Omtrek berekenen

Exercise type: Creative
Maximum points: 1

THEORIE: Lees de uitleg goed!

De belangrijkste **eenheden** van **lengte** zijn:

kilometer = km = 1000 meter
 hectometer = hm = 100 meter
 decameter = dam = 10 meter
 meter = m = 1 meter
 decimeter = dm = 0,1 meter
 centimeter = cm = 0,01 meter
 millimeter = mm = 0,001 meter

Met onderstaand schema kan je **lengtematen omrekenen!**
 Van links naar rechts krijg je **per stapje** x10.
 Van rechts naar links krijg je **per stapje** :10.

Probeer dit schema bij elke vraag te gebruiken.

km	hm	dam	m	dm	cm	mm	
x 10		x 10		x 10		x 10	
: 10		: 10		: 10		: 10	

Combineren maar!

Exercise type: Match pairs
Maximum points: 5

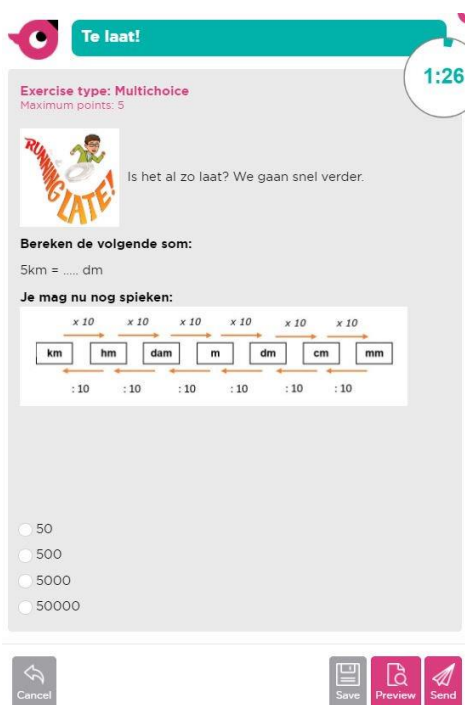
1:17

Verbind de woorden met de cijfers.

Hecto	0,01
Kilo	10
Deca	1000
Centi	0,001
Mili	100

Buttons: Cancel, Save, Preview, Send

Figuur 5. Theorieblok Eenheden van Lengte. Figuur 6. Sleepvraag.



Figuur 7. Meerkeuzevraag.

De keuze voor variatie in opdrachten is gemaakt omdat leerlingen deze behoefte aangaven bij een eerdere game van de onderzoeker. Leerlingen gaven hierbij aan dat ze alleen meerkeuzevragen als saai ervoeren. Naast kennisvragen zijn ook opdrachten toegevoegd waarbij spelers de instructie nogmaals aan zichzelf uitleggen (= self-explanation). Deze vragen zijn voor spelers ook herkenbaar door een steeds terugkerende avatar (Figuur 8). Na het uitspelen van MeTriek krijgt de speler het totaal aantal behaalde punten en vuurwerk met een beker te zien.



Figuur 8. Avatar self-explanation.

Voorafgaand aan de uitvoering van de pilot is door de onderzoeker contact opgenomen met de schoolleider van de vmbo-school voor toestemming. Na toestemming van de schoolleider is volgens cETO-richtlijnen (Research Ethics Committee van de Open Universiteit) de werving gestart. In

verband met de privacy van de leerlingen hebben de mentoren de ouders en leerlingen via een schriftelijk *informed consent* toestemming gevraagd voor deelname aan de pilot. Leerlingen en ouders konden deze informed consent twee weken inleveren bij de mentor. De leerlingen met toestemming zijn door de mentoren van de school samengevoegd op een totaallijst en zijn op basis van hun wiskundecijfers op numerieke volgorde gezet van hoog naar laag. Uit beide klassen werd de leerling met het hoogste, middelste (mits voldoende) en laagste cijfer gekozen om deel te nemen aan de pilot. De onderzoeker ontving alleen een envelop met de toestemmingsverklaringen van deze zes leerlingen. Omdat de onderzoeker met anonieme inlogcodes werkt zijn de namen van de leerlingen ook niet aan de onderzoeker bekend gemaakt. De leerlingen die niet gekozen zijn konden in overleg met de mentor MeTriek spelen, waarbij geen gegevens verzameld werden door de onderzoeker. MeTriek was via <https://play.seppo.io/> bereikbaar voor de leerlingen. Leerlingen konden in het klaslokaal met een unieke code en een laptop van school aan de slag. Tijdens de cognitive walkthrough speelde steeds één leerling MeTriek in aanwezigheid van de onderzoeker. Daarna vulde de leerling de vragenlijst in die de educatieve kwaliteit van MeTriek meet. Na het invullen van deze vragenlijst zijn de antwoorden individueel met de onderzoeker besproken, zodat de onderzoeker toelichting kon vragen bij antwoorden. Alleen als vragen onduidelijk waren kreeg de leerling hulp tijdens het invullen van de vragenlijst.

De evaluatie van het prototype liet tijdens de cognitive walkthrough zien dat 4 van de 6 leerlingen niet begrepen dat ze na elke instructie de instructie zichzelf nog een keer moesten uitleggen. Bij navragen bleek dat dit kwam omdat het theorieblok en de vraag in één tekstblok stond. De veronderstelling van de onderzoeker was dat MeTriek binnen 10 á 20 minuten uit te spelen was met het feedbacktype EF. Dit bleek voor alle leerlingen zo te zijn. Uit de informatie verkregen uit het nagesprek bleek dat leerlingen het te lastig vonden dat vijf verschillende onderwerpen aan bod kwamen (één onderwerp per level). Ze gaven in de gesprekken aan dat ze hierdoor te veel verschillende stof moesten leren. Tijdens het invullen van de vragenlijst is gebleken dat een aantal vragen te moeilijk waren voor vmbo-leerlingen. Ze gaven aan sommige zinnen onduidelijk te vinden zoals: 'Hoe mentaal belastend waren de opdrachten in de game'. Uit de analyse van de antwoorden op de vragenlijsten bleek dat MeTriek op alle onderdelen voldoende scoorde. De score is berekend door per categorie een gestandaardiseerde score te berekenen, waarbij (>.50) voldoende was. De scores van de game-versie tijdens de Pilot waren: identity (.67), immersion (.64), interaction (.66), increased complexity (.59), informed teaching (.69) en instructional (.53).

Aan de hand van deze resultaten is besloten om van vijf naar twee onderwerpen te gaan, om op deze manier de game minder lastig te maken waardoor meer opdrachten over hetzelfde onderwerp konden worden opgenomen. Het was daarbij wel belangrijk dat genoeg variatie in vragen mogelijk zou blijven. Daarom heeft de onderzoeker gekozen voor twee grotere onderwerpen, namelijk het rekenen met lengtematen en het berekenen van de omtrek. Door minder onderwerpen te gebruiken verwachtte de onderzoeker de complexiteit van de game meer in overeenstemming te brengen met de

behoefte van de leerling, en hiermee de score op increased complexity te verhogen. Het overzien van de self-explanation vragen is opgelost door een theorieblok aan te bieden en te vragen of het theorieblok is gelezen. De eigen uitleg door de speler wordt nu bij de volgende vraag gevraagd. De leerling ziet door deze scheiding vermoedelijk eerder het verschil tussen een theorieblok en een vraag. De onderzoeker verwachtte door deze aanpassing dat het onderdeel instructional in de game tijdens het interventieonderzoek beter zou zijn dan bij de pilot. Om de leesbaarheid van de vragenlijst te verbeteren is gekozen om een (klein) aantal vragen te herformuleren. Dit leidde tot een aangepaste vragenlijst voor de meting van de educatieve kwaliteit van MeTriek bij het interventieonderzoek (zie Bijlage 2). Hierna is beschreven hoe het interventieonderzoek is uitgevoerd met deze bijgestelde versie voor de feedbacktype-specifieke varianten van MeTriek.

3. Methode

3.1 Ontwerp

Het interventieonderzoek gebruikt een gerandomiseerd pretoets-posttoets-retentietoets ontwerp naar de leereffecten van feedbacktype KR, KCR en EF op het wiskundedomein lengtematen en het berekenen van de omtrek. Er zijn twee experimentele groepen (feedbacktypes KCR en EF) en één controlegroep (feedbacktype KR).

De leerlingen zijn aselekt over deze drie onderzoeksgroepen verdeeld en kregen tijdens elke toets (de pretoets, posttoets en retentietoets) twaalf vragen om via de leerresultaten de mate van beheersing van het wiskundeonderwerp te kunnen meten. De mate van beheersing van dit wiskundeonderwerp is de afhankelijke variabele. De onafhankelijke variabele is het feedbacktype. Na de pretoets speelden de leerlingen MeTriek, waarbij ze bij elke opdracht de feedbacktype-specifieke feedback kregen op de uitwerking van die opdracht. Na het spelen kregen leerlingen de posttoets om het leereffect van het geven van feedback op korte termijn te bepalen, namelijk via het meten van leerprestaties van de leerlingen op de posttoets. Twee weken ná het spelen van de MeTriek is de retentietoets afgenomen. De retentietoets meet het leereffect van de feedbacktypes op langere termijn.

3.2 Participanten

Alle leerlingen (N = 161) van een vmbo school in Noord-Brabant zijn gevraagd om deel te nemen aan het interventieonderzoek. Van deze leerlingen deden 74 leerlingen de vmbo Basis Beroepsgerichte Leerweg (BBL), 49 leerlingen de Kader Beroepsgerichte Leerweg (KBL), en 38 leerlingen de Gemengde Leerweg (GL). Het wiskundeonderwerp uit dit onderzoek wordt bij elke Leerweg in het eerste jaar tijdens de lessen behandeld.

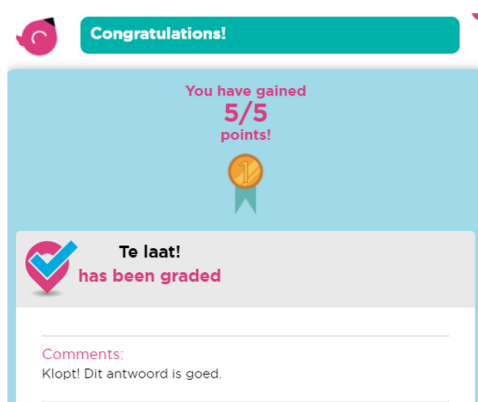
Uitval en een lage startrespons waren bedreigingen voor de validiteit en de betrouwbaarheid van dit onderzoek. Het feit dat de wiskundeleraars enthousiast waren over het onderzoek hielp bij het tegengaan van een lage startrespons. De wiskundeleraars hebben meerdere keren het onderzoek

gepromoot in de klas. Door de Covid-19 pandemie was het niet mogelijk om als onderzoeker voorafgaand aan de interventie te gaan promoten in de klassen. De onderzoeker heeft in plaats daarvan een video gemaakt om ouders en leerlingen te enthousiasmeren. Deze video beoogde een lagere startrespons te voorkomen vanwege de Covid-19 maatregelen. Daarnaast werd verwacht dat er meer dan gebruikelijke uitval zou zijn door de Covid-19 maatregelen, omdat de metingen op school moesten plaatsvinden.

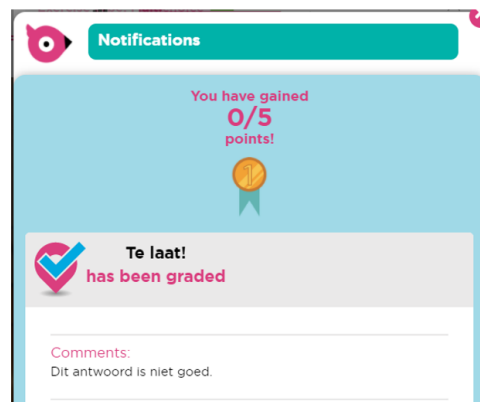
Nadat de onderzoeker en de mentoren herhaalde verzoeken tot deelname deden, wilden 130 leerlingen aan het onderzoek deelnemen. Omdat niet alle leerlingen door de Covid-19 maatregelen bij alle meetmomenten aanwezig konden zijn is een deel van deze leerlingen voor het onderzoek uitgevallen. Daarnaast heeft één klas, waarvan 6 leerlingen zich hadden opgegeven, wegens omstandigheden niet deelgenomen aan het onderzoek. Uiteindelijk hebben 105 leerlingen aan alle metingen deelgenomen. Dit is een responspercentage van 80.8%. Dit aantal is groter dan het minimaal benodigde aantal van 81 participanten dat gevonden is op basis van een berekening in G*Power waarbij uitgegaan is van een effectgrootte van $d = .40$ (Faul, Erdfelder, Lang, & Buchner, 2007). van der Kleij et al. (2015) hebben in hun meta-analyse voor elaboratieve feedback een effectgrootte van $d = .49$ gevonden en voor wiskunde-games waren de effectgroottes zelfs $d = .93$. van der Kleij et al. (2015) constateerden dat nauwelijks onderzoeken zijn gedaan bij wiskunde games in combinatie met elaboratieve feedback, daarnaast verschilde de operationalisatie van elaboratieve feedback sterk. In deze scriptie is daarom uitgegaan van een lagere effectgrootte voor het feedbacktype EF.

3.3 Materialen

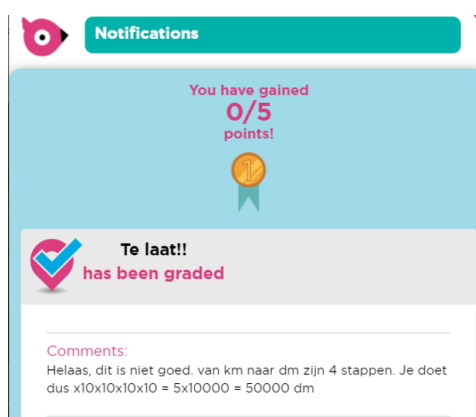
De wiskundegame MeTriek is ná de Pilot aangepast (zie Pilot MeTriek) en is uitgewerkt voor de drie feedbacktype-specifieke varianten in het interventieonderzoek. De leerlingen krijgen een ander feedbacktype na het beantwoorden van een vraag, afhankelijk van de groep waarin ze zijn ingedeeld (Figuur 9 t/m 14).



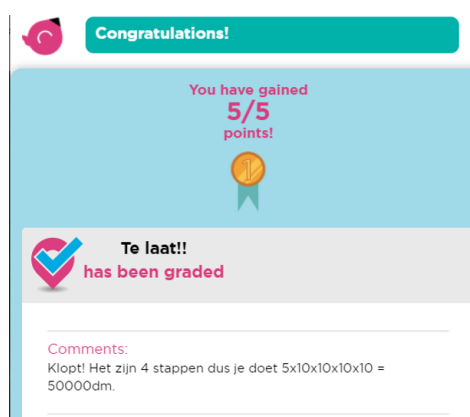
Figuur 9. Feedbacktype KR - Goed



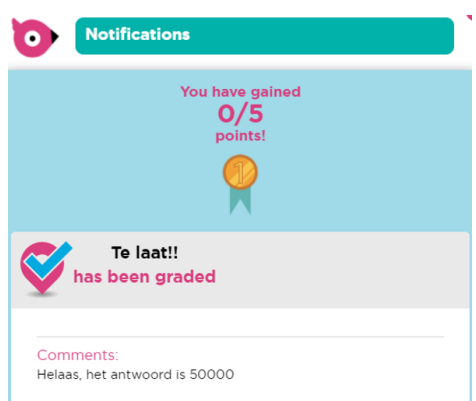
Figuur 10. Feedbacktype KR - Fout



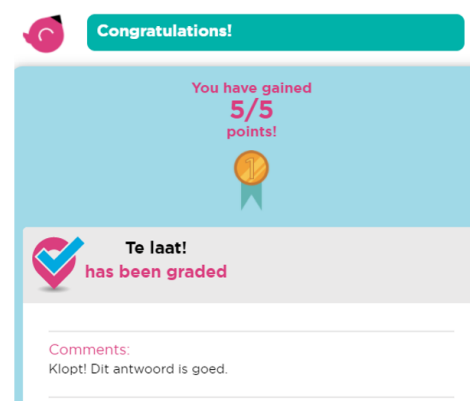
Figuur 11. Feedbacktype EF – Fout



Figuur 12. Feedbacktype EF – Goed



Figuur 13. Feedbacktype KCR – Fout



Figuur 14. Feedbacktype KCR – Goed

3.4 Meetinstrumenten

Tijdens het interventieonderzoek zijn drie toetsen gebruikt om de beheersing van het rekenen met lengtematen en omtrek te meten (Bijlage 4). Elke toets bestaat uit 12 vragen. De pre-, post- en retentietoets zijn door de onderzoeker samengesteld met toetsvragen naar analogie van de toetsvragen die gebruikt worden op de school waarin het onderzoek werd uitgevoerd. Hierdoor is de inhoudsvaliditeit van deze toetsen gegarandeerd. Om de moeilijkheidsgraad van de drie toetsen gelijk te houden zijn de vragen van de drie toetsen tegelijkertijd opgesteld. Eerst is voor elke toets vraag één opgesteld en daarna vraag twee etc. De vragen zijn in opzet hetzelfde, maar de context en/of de getallen die gebruikt zijn verschillen van elkaar. Bovendien zijn de toetsen door een andere wiskundedocent gecontroleerd op vergelijkbare moeilijkheidsgraad en representativiteit voor dit wiskundeonderwerp. Door het gebruik van meerkeuzevragen en open vragen waarbij maar één goed antwoord mogelijk is, bestaat geen twijfel over de bepaling van de scores op de toetsen.

Naast de scores van de kennistoetsen werden in Seppo per speler ook de gamescores, en de sessieduur in uren, minuten en seconden bijgehouden. De veronderstelling is dat leerlingen wellicht meer tijd besteden aan de feedback bij het feedbacktype EF dan bij de andere feedbacktypes (Timmers

& Veldkamp, 2011). De feedback bij variant EF heeft namelijk iets meer tekst en daarom leestijd. De vraag is echter of de sessieduur bij de leerlingen met het feedbacktype EF langer is dan bij de andere feedbacktypes.

Tot slot werd verondersteld dat leerlingen met het feedbacktype EF hogere gamescores hebben omdat ze bij het lezen van de uitgebreidere feedback intensiever bezig zijn met de leerstof en door deze uitleg een volgende opdracht beter zullen maken.

Om de educatieve kwaliteit van MeTriek te meten op de zes elementen van Annetta (2010) is gebruik gemaakt van de aangepaste vragenlijst van Kouwenhoven (2019) (Bijlage 2). Deze vragenlijst is in dat onderzoek voldoende intern consistent gebleken (Kouwenhoven, 2019).

3.5 Procedure

Voorafgaand aan het interventieonderzoek is de onderzoeker in contact gekomen met de coördinator externe contacten van de vmbo school voor het interventieonderzoek. Het interventieonderzoek zou eind maart 2020 uitgevoerd worden. Helaas heeft dit door de Covid-19 pandemie niet plaats kunnen vinden en is het interventieonderzoek acht maanden uitgesteld. Het contact met de coördinator van de school voor het interventieonderzoek is altijd gebleven, waardoor samen werd besloten om in november te starten met een nieuw onderwerp, waardoor een nieuwe game gemaakt moest worden.

Ouders en leerlingen hadden vijf weken de tijd om door middel van een schriftelijk of online informed consent toestemming te geven voor de deelname aan het interventieonderzoek.

Leerlingen en ouders konden deze informed consent op papier inleveren bij de wiskundedocent of mailen naar de onderzoeker. Vanwege coronamaatregelen was het mogelijk om via een online informed consent toestemming geven. Na twee en vier weken is door de onderzoeksschool een herinnering verstuurd naar de ouders en leerlingen. Leerlingen en ouders hadden altijd de mogelijkheid om af te zien van deelname aan het onderzoek. Leerlingen die niet deelnamen aan het onderzoek of zich tijdens het onderzoek terug zouden trekken speelden wel MeTriek en maakten ook de kennistoetsen omdat de lesstof bij het schoolprogramma hoorde. Voor deze leerlingen zijn echter unieke inlogcodes aangemaakt binnen een andere versie van MeTriek. Deze versie is direct na het spelen van de Seppo server verwijderd en de resultaten op de kennistoetsen zijn niet gebruikt voor het onderzoek.

Alle leerlingen uit de verschillende vmbo-klassen die toestemming gaven voor deelname zijn per leerniveau (= onderwijsniveau, het vmbo-niveau waar de leerling zit) op een lijst gezet om daarna op alfabetische volgorde in één van de drie onderzoeksgroepen (EF, KCR, KR) ingedeeld te worden. De leerlingen op de 'vmbo-BBL' lijst werden eerst verdeeld in onderzoeksgroepen, daarna de vmbo-KBL leerlingen en daarna de vmbo-GL leerlingen. Hier is voor gekozen om de leerlingen met verschillende leerniveaus te spreiden over de verschillende groepen, waardoor een onderzoeksgroep gemiddeld hetzelfde leerniveau had.

De eerste leerling werd ingedeeld in onderzoeksgroep EF, de tweede leerling in onderzoeksgroep KCR, de derde leerling in onderzoeksgroep KR, de vierde leerling in onderzoeksgroep EF etc. De leerlingen kregen een individuele code toegekend voor het onderzoek. Deze code bestond uit A (bij EF), B (bij KCR) of C (bij KR) met een combinatie van cijfers. Welke code gekoppeld was aan welke leerling was alleen bij de onderzoeker bekend vanwege de privacy van de leerlingen. De onderzoeker zorgde ervoor dat de wiskundedocent een pakket had met de juiste toetsen per onderzoeksgroep, gekoppeld aan de persoonlijke codes van de leerling.

De pretoets vond plaats in de eerste week van het onderzoek. De leerlingen uit de controlegroep (KR) en de twee experimentele groepen (KCR en EF) maakten tegelijkertijd de toets in het lokaal met de wiskundedocent, meestal onder toezicht van de onderzoeker. In dezelfde week als de pretoets-afname speelden de leerlingen de versie van MeTriek waaraan ze waren toegewezen met de individuele code. Leerlingen speelden in het eigen klaslokaal onder toezicht van de wiskundedocent en (soms) de onderzoeker. Voor het spel kreeg elke leerling een unieke code, waarmee hij/zij gekoppeld was aan de juiste feedback specifieke variant. De leerlingen kregen tijdens de les maximaal 30 minuten de tijd om MeTriek te spelen. De wiskundedocenten en de onderzoeker zagen erop toe dat leerlingen tijdens het spelen niet met elkaar konden overleggen.

Direct ná het spelen van MeTriek beantwoordden de leerlingen in dit blok de vragen uit de posttoets en vulden via LimeSurvey de online vragenlijst in over de educatieve kwaliteit van gebruikte variant van MeTriek. De retentietoets heeft twee weken later plaatsgevonden. De drie verschillende toetsen zijn nagekeken door de onderzoeker. Een grafische weergave van de tijdlijn voor het interventieonderzoek staat weergegeven in Figuur 15.

Interventieonderzoek									
Werving respondenten	Werving respondenten	Werving respondenten	Werving respondenten	Werving respondenten	Introductie MeTriek en onderzoek	Pretoets, interventie, posttoets, vragenlijst Educatieve kwaliteit	Nakijken toetsen	Nakijken toetsen	Retentietoets
Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14

Figuur 15. Verloop Interventieonderzoek

3.6 Data-analyse

Om de data te analyseren is gebruik gemaakt van IBM SPSS 27 statistics. Met behulp van de Shapiro-Wilk test is gekeken of de samengenomen data van de onderzoeksgroepen normaal verdeeld waren. Ter aanvulling is ook met Q-Q plots, histogrammen en boxplots naar de normaalverdeling gekeken. De grafieken van de samengenomen data van de onderzoeksgroepen en de data ingedeeld op feedbacktypes zag er normaal uit waardoor is voldaan aan de assumpties voor het toepassen van een ANOVA test. Vanwege de uitval van respondenten na de start van het onderzoek is een Chi-kwadraattoets gebruikt om te kijken of de verdeling van de leerniveaus (BBL, KBL en GL) binnen de

onderzoeksgroepen niet te veel van elkaar verschilden. In de Chi-kwadraattoets is ook gekeken naar de verdeling tussen het aantal jongens en meisjes, omdat onderzoek uitwijst dat veel verschillen bestaan tussen gamen onder jongens en meisjes (van Rooij, Dalinghaus, & van den Eijnden, 2018). Verder is met een One-Way ANOVA gekeken naar de verschillen in pretoets-scores, gamescores, sessieduur en leeftijd.

In het ontwerp van dit onderzoek is één between-subject variabele opgenomen (feedbacktype) en één within-subject variabele (tijd, waarbij de scores van de pretoets, posttoets en retentietoets zijn ingevoerd). De toetsing vond plaats aan de hand van een mixed ANOVA met repeated measures, met een significantieniveau van $p \leq .05$. De effectgroottes r zijn handmatig berekend, omdat dit niet mogelijk is in SPSS. Volgens Field (2013) is r een goede effectgrootte om te gebruiken bij een (mixed) ANOVA met repeated measures.

Met een betrouwbaarheidsanalyse is de betrouwbaarheid van de drie kennistoetsen geanalyseerd.

Voor de bepaling van de educatieve kwaliteit van de serious game MeTriek zijn de data eerst vanuit LimeSurvey in SPSS gezet. Een aantal vragen zijn omgepoold waarna vervolgens betrouwbaarheidsanalyses in SPSS zijn uitgevoerd op de zes elementen van Annetta (2010). Tot slot is de educatieve kwaliteit van MeTriek in Excel bepaald door gestandaardiseerde scores voor deze zes elementen te berekenen (tussen 0 en 1). Een element werd als voldoende beoordeeld als de gemiddelde score $> .5$ was en was onvoldoende bij een gemiddelde score $< .5$.

4. Resultaten

4.1 Beschrijvende statistieken

Het responspercentage in het interventieonderzoek was 80.8% ($N = 105$). In Tabel 1 staan de algemene gegevens over het niveau, het geslacht en de leeftijd van de respondenten. In de onderzoeksgroepen blijkt geen significant verschil te zitten in de verdeling tussen jongens en meisjes $\chi^2(2) = 1.68, p = .43$. Verder blijkt de gemiddelde leeftijd bij de verschillende onderzoeksgroepen niet significant van elkaar te verschillen $F(2, 102) = .10, p = .91$. De verdeling van de respondenten over de drie onderwijsniveaus (BBL, KBL, GL) is binnen elke onderzoeksgroep vergelijkbaar $\chi^2(4) = 0.47, p = 1.00$.

Tabel 1

Algemene gegevens respondenten

	N	Niveau			Geslacht		Leeftijd
		BBL	KBL	GL	Man	Vrouw	M (SD)
EF	39	14	13	12	19	20	12.23 (.54)
KCR	34	12	11	11	18	16	12.29 (.76)
KR	32	11	11	10	12	20	12.25 (.57)
Totaal	105	37	35	33	49	56	12.26 (.62)

Noot. Leeftijd in jaren.

4.2 Leerresultaten

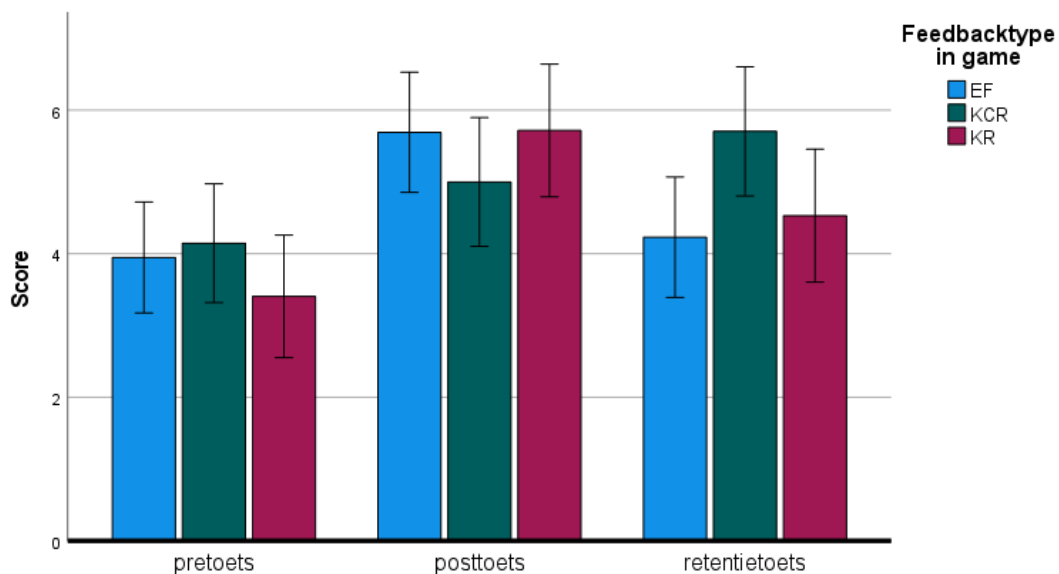
Om de betrouwbaarheid van de kennistoetsen te meten is een geaggreerde betrouwbaarheidsanalyse uitgevoerd, waarbij de deelnemers van de onderzoeksgroepen EF, KCR en KR zijn samengevoegd ($N = 105$). Uit de betrouwbaarheidsanalyse blijkt dat de betrouwbaarheid van toets A (12 items, $\alpha = .75$) en B (12 items, $\alpha = .72$) acceptabel zijn en toets C twijfelachtig is (12 items, $\alpha = .68$).

De leerresultaten werden bepaald via de toetsscores. De toetsscores op de pretoets, de posttoets en de retentietoets staan in Tabel 2 en een grafische weergave in Figuur 16. Eerst is met een One-way ANOVA gekeken of de scores van de pretoets tussen de onderzoeksgroepen significant van elkaar verschilden. Uit de resultaten bleek dat geen significante verschillen zijn tussen de scores bij de pretoets $F(2, 102) = .82, p = .44$. De onderzoeksgroepen verschilden dus - zoals op basis van de randomisering van de onderzoeksgroepen werd verwacht - niet op voorkennis voor het wiskundeonderwerp in MeTriek.

Tabel 2

Gemiddelde scores op pretoets, posttoets en retentietoets per onderzoeksgroep

	Feedbacktype	N	M	SD	SE	CI (95%)
Pretoets	EF	39	3.95	2.60	.42	[3.11-4.79]
	[0-12]					
	KCR	34	4.15	2.41	.41	[3.30-4.99]
	KR	32	3.41	2.24	.40	[2.60-4.21]
	Totaal	105	3.85	2.43	.24	[3.38-4.32]
Posttoets	EF	39	5.69	2.67	.43	[4.83-6.56]
	[0-12]					
	KCR	34	5.00	2.94	.51	[3.97-6.03]
	KR	32	5.72	2.22	.39	[4.92-6.52]
	Totaal	105	5.48	2.63	.26	[4.97-5.99]
Retentietoets	EF	39	4.23	2.63	.36	[3.51-4.95]
	[0-12]					
	KCR	34	5.71	3.11	.53	[4.62-6.79]
	KR	32	4.53	2.58	.46	[3.60-5.46]
	Totaal	105	4.80	2.69	.46	[3.60-5.46]



Figuur 16. Grafische weergave van de gemiddelde scores op de pretoets, posttoets en retentietoets met een spreiding van CI (95%).

Met een mixed ANOVA met repeated measures op de scores van de pretoets, posttoets en retentietoets zijn de hypothesen getoetst. Allereerst is nagegaan of het spelen van MeTriek voor de controlegroep (KR) en experimentele groepen (EF, KCR) leidt tot een verbetering van de leerresultaten over de tijd gemeten (hoofdeffect). Daarna is het hoofddoel van het interventieonderzoek nagegaan, namelijk het vergelijken van de leerresultaten van de drie feedbacktypes per meetmoment. Hierbij is enerzijds nagegaan of het geven van EF feedback zorgt

hogere leerresultaten op de posttoets en de retentietoets ten opzichte van KCR feedback en KR feedback. Anderzijds is nagegaan of het geven van KCR feedback zorgt voor hogere leerresultaten op de posttoets en de retentietoets ten opzichte van KR feedback.

Mauchly's test laat zien dat is voldaan aan de aanname van sphericiteit $X(2) = .99, p = .68$. De resultaten laten zien dat de gemiddelde toetsscores bij het samennemen van de onderzoeksgroepen over tijd significant van elkaar verschilden bij zowel de posttoetsscores als de retentietoetsscores; hoofdeffect van tijd, $F(2, 204) = 33.14, p < .05$. Het contrast (*Simple*), waarbij de pretoetsscores met de posttoetsscores werden vergeleken en de pretoetsscores met de retentietoetsscores, liet zien dat de onderzoeksgroepen significant hoger scoorden op de posttoets na het spelen van MeTriek $F(1, 102) = 60.27, p < .05$. Naast deze hogere score op korte termijn scoorden de samengenomen onderzoeksgroepen ook significant hoger op de retentietoets dan op de pretoets $F(1, 102) = 25.37, p < .05$.

Uit de Test of Within-Subjects Effects bleek een significant interactie-effect tussen de leerresultaten over tijd en de feedbacktypes bij de samengenomen onderzoeksgroepen $F(4, 204) = 6.41, p < .05$. Uit het contrast (*Simple*) bleek dat de samengenomen onderzoeksgroepen significant hoger scoorden op de posttoets dan op de pretoets $F(1, 102) = 3.88, p < .05$ en dat de samengenomen onderzoeksgroepen ook op de retentietoets een hogere score behaalden dan op de pretoets $F(2, 102) = 3.86, p < .05$. Uit deze analyses bleek dat MeTriek voor de samengenomen onderzoeksgroepen een significant positief effect had op de leerresultaten, zowel op de korte termijn, als ook op de lange termijn.

Om te zien of dit ook geldt voor elk van de afzonderlijke onderzoeksgroepen (EF, KCR en KR) is gekeken naar Post-hoc toetsen. Hierna worden alleen significante effecten vermeld. De Post-hoc toetsen tonen aan dat bij feedbacktype EF op de posttoets gemiddeld 1.74 hoger werd gescoord dan op de pretoets ($p < .05, r = .64$). Bij feedbacktype KCR is gemiddeld 1.56 hoger gescoord op de retentietoets dan op de pretoets ($p < .05, r = .56$). Bij feedbacktype KR is op de posttoets gemiddeld 2.31 hoger gescoord dan op de pretoets ($p < .05, r = .80$) en gemiddeld 1.13 hoger op de retentietoets dan op de pretoets ($p < .05, r = .54$). Hypothese 1 wordt deels bevestigd, waaruit de potentie van Metriek voor leren blijkt.

Om het hoofddoel van het interventieonderzoek na te gaan is gekeken naar de resultaten van de Tests of Between-Subjects Effect. Uit deze test bleek geen significante verschillen te zijn tussen de verschillende feedbacktypes $F(2, 102) = .30, p = .75$. Met een *Simple Effect* analyse is gekeken of verschillen zijn tussen de drie feedbacktypes op de scores voor de posttoets en retentietoets. Uit deze analyse bleek dat bij de posttest geen significante verschillen tussen de drie onderzoeksgroepen op hun scores waren $F(2, 102) = .82, p = .44$. en bij de retentietoets wel $F(2, 102) = 3.07, p = .05$. Uit de pairwise comparisons bleek alleen een significant verschil tussen feedbacktype EF en KCR op hun scores voor de retentietoets. Onderzoeksgroep EF scoorde daar gemiddeld 1.48 lager dan onderzoeksgroep KCR ($p < .05$). Hypotheses 2a en 2b worden verworpen omdat de drie feedbacktypes

onderling vrijwel altijd niet significant te verschillen op hun leereffecten (op een bepaald moment en door de tijd).

4.3 Gamescores en sessieduur

De onderzoeksgroepen verschillen toch niet in sessieduur $F(2, 102) = 1.7, p = .19$ en ook niet op de behaalde gamescore in MeTriek $F(2, 102) = 1.4, p = .25$. (zie Tabel 3).

Tabel 3

Sessieduur en gamescores

Feedbacktype	Sessieduur					Gamescores		
	N	M	SD	Min	Max	M (SD)	Min	Max
EF	39	16.6	5.26	7.5	29.2	61.56 (10.51)	36	87
KCR	34	17.9	5.07	10.8	27.9	58.62 (14.65)	8	83
KR	32	15.7	5.44	6.8	25.3	64.63 (14.55)	36	90
Totaal	105	16.8	5.28	6.8	25.3	61.54 (13.32)	8	90

Noot. Sessieduur is gemeten in minuten, de maximale gamescore was 99.

4.4 Educatieve kwaliteit

Voor de meting van de educatieve kwaliteit van MeTriek is de licht aangepaste vragenlijst van Kouwenhoven (2019) gebruikt (Bijlage 2). De betrouwbaarheid van deze vragenlijst is gemeten door Cronbach's alpha (α) te berekenen. Eerst is een geaggreerde betrouwbaarheidsanalyse uitgevoerd (EF, KCR, KR). Uit deze analyse blijkt dat de interne consistentie bij de subschaal identity twijfelachtig is (2 items, $\alpha = .69$). Immersion heeft een uitstekende consistentie (26 items, $\alpha = .93$). De subschaal interactivity bleek acceptabel (2 items, $\alpha = .70$). Complexity bleek slecht betrouwbaar (8 items, $\alpha = .58$), informed teaching bleek twijfelachtig (3 items, $\alpha = .63$) en instructional acceptabel (2 items, $\alpha = .71$).

In Tabel 4 staan de resultaten van de kwaliteitsmeting voor de game MeTriek. MeTriek scoorde bij het samenvoegen van de onderzoeksgroepen en per onderzoeksgroep voldoende op de zes elementen van Annetta (2010).

Tabel 4

Scores educatieve kwaliteit MeTriek

	Identity	Immersion	Interactivity	Increasing complexity	Informed teaching	Instructional
EF	.64	.68	.68	.69	.65	.73
KCR	.64	.66	.72	.64	.69	.73
KR	.64	.70	.73	.65	.68	.74
Totaal	.64	.68	.68	.66	.66	.77

Noot. Gemiddelde scores waarbij >.5 voldoende is.

5. Discussie en conclusie

Deze scriptie staat vooral in het teken van het interventieonderzoek naar het leereffect van drie feedbacktypes met de feedbacktype-specifieke varianten van de zelfontwikkelde serious game MeTriek. Om voldoende zeker te zijn dat MeTriek afdoende educatieve kwaliteit heeft om van te kunnen leren, is in het eerste deel van deze scriptie met een pilot de educatieve kwaliteit bepaald van MeTriek. MeTriek is ontworpen op basis van het framework van Annetta (2010). De toepassing van dit framework richt zich op de borging van de educatieve kwaliteit van een serious game. In de pilot bleek dat MeTriek al voldoende scoorde op educatieve kwaliteit, gemeten met het gevalideerde instrument van Kouwenhoven (2019), maar dat verbeteringen mogelijk waren die zijn doorgevoerd in een aangepaste versie van MeTriek voorafgaand aan het interventieonderzoek, dat is uitgevoerd in het tweede deel van deze scriptie. Dit interventieonderzoek met een gerandomiseerd pretoets-posttoets-retentietoets ontwerp had als doel meer kennis te verwerven over de effectiviteit van verschillende feedbacktypes op de leerresultaten in serious games voor het vmbo-onderwijs, hier toegepast voor MeTriek, een game gericht op het leren berekenen van lengtematen en omtrek. Uit de kwaliteitsmeting bij het interventieonderzoek bleek dat deze aangepaste versie van MeTriek voor alle feedbacktypes (EF, KCR en KR) van voldoende educatieve kwaliteit was. De MeTriek-versie tijdens het interventieonderzoek scoorde in vergelijking met de pilot-versie hoger op immersion, interactivity, complexity, informed teaching en instructional.

De centrale vraag in het interventieonderzoek was: “Wat is het leereffect van de feedbacktypes KR, KCR en EF bij het spelen van de serious game MeTriek door leerlingen in het vmbo?” Hierbij is KR de controlegroep en zijn EF en KCR de experimentele groepen. Vooraf werden de volgende hypothesen opgesteld:

- Hypothese 1: Het spelen van de serious game MeTriek leidt bij elk feedbacktype tot een positief leerresultaat.

- Hypothese 2a: Het spelen van de serious game MeTriek met feedbacktype EF leidt tot betere leerresultaten dan het spelen van MeTriek met feedbacktype KCR en KR.
- Hypothese 2b: Het spelen van de serious game MeTriek met feedbacktype KCR leidt tot betere leerresultaten dan het spelen van MeTriek met feedbacktype KR.

De scores op de pretoets bevestigen de verwachting dat de onderzoeksgroepen door hun randomisering niet verschillen in voorkennis op het wiskundeonderwerp dat met het spelen van MeTriek geleerd kan worden. Het spelen van MeTriek toont – voor het feedbacktype KR - verbeterde leerresultaten, zowel op de korte termijn (pretoets versus posttoets) als op de lange termijn (pretoets versus retentietoets). Het spelen van MeTriek toont voor het feedbacktype KCR verbeterde leerresultaten op de lange termijn en voor het feedbacktype EF alleen verbeterde leerresultaten op de korte termijn. De onderzoeksgroepen hadden een vergelijkbare tijdsbesteding. Hypothese 1 wordt deels bevestigd, waaruit de potentie van Metriek voor leren blijkt.

De drie feedbacktypes blijken onderling vrijwel altijd niet significant te verschillen op hun leereffecten (op een bepaald moment en door de tijd). Hypotheses 2a en 2b worden daarom verworpen.

5.1 Vergelijking met eerder onderzoek

De resultaten van dit onderzoek zijn niet in lijn met eerder onderzoek (Shute, 2008; Butler et al., 2013; van der Kleij et al., 2015). Verwacht werd dat EF bij het vak wiskunde een groot leereffect zou hebben en effectiever zou zijn dan KCR en KR. In het huidige onderzoek werden echter tussen de drie gebruikte feedbacktypes geen verschillen gevonden in hun leereffecten.

5.2 Sterkte punten en beperkingen in het huidige onderzoek

Dit onderzoek kent zowel sterke punten als beperkingen. Door de pilot kon het ecologisch valide interventieonderzoek in een beter bekende setting worden uitgevoerd en was zorg gedragen voor voldoende educatieve kwaliteit van MeTriek. Ecologisch valide onderzoek is minder controleerbaar, maar heeft als sterk punt dat de bevindingen eenvoudig naar de praktijk vertaald kunnen worden. De randomisatie in het interventieonderzoek is eveneens een sterk punt. In dit onderzoek zijn deelnemers van zeven klassen op leerniveau aselekt over de drie onderzoeksgroepen verdeeld. Hierbij bleven de leerlingen tijdens het onderzoek in hun eigen vertrouwde klas. Dit had als organisatorisch voordeel dat het lesrooster behouden kon blijven en dat alle leerlingen uit één klas onder toezicht stonden van dezelfde docent. Door dit toezicht ontstond geen mogelijkheid voor overleg tijdens het spelen van MeTriek en tijdens het maken van de toetsen. Een ander voordeel van het interventieonderzoek was dat de onderzoeker geprobeerd heeft om de interventie zoveel mogelijk tegelijk te laten plaatsvinden en de posttoets direct ná het spelen van MeTriek af te nemen, waardoor dat klasoverstijgend overleg nauwelijks mogelijk was, wat de kans op *diffusion of treatments* verkleint (Cresswell, 2014).

In het interventieonderzoek is een pretoets gebruikt, wat volgens Cresswell (2014) het voordeel heeft dat de onderzoeker inzicht krijgt in het startniveau van de leerlingen en de vergelijkbaarheid.

Andere (praktische) voordelen waren dat leerlingen allemaal beschikten over een eigen laptop, waardoor dat ze al bekend waren met het eigen materiaal en zonder (veel) uitleg aan de slag konden met MeTriek. Naast de leerlingen, hebben enthousiaste wiskundedocenten er bij dit onderzoek voor gezorgd dat de werving (zelfs) ondanks de Covid-19 pandemie soepel is verlopen, waardoor voldoende respondenten deelnamen aan dit onderzoek.

Het uitvoeren van dit interventieonderzoek in een ecologisch valide setting geeft een duidelijker beeld op de praktische uitvoerbaarheid dan een gecontroleerd experiment, maar heeft ook een vijftal beperkingen die de bevindingen uit dit onderzoek van een aantal kanttekeningen voorzien.

Ten eerste was de voorkennis hoger dan verwacht, waardoor beperktere ruimte voor leerwinst ontstond. Ten tweede was slechts beperkte leerwinst mogelijk omdat in het onderzoek minder tijd (ca. 25 min) aan het onderwerp besteed werd dan anders tijdens de reguliere lessen (ca. 180-220 min). Omdat het onderwerp anders over meerdere lessen (sessies) is gespreid, is vanwege de betere controleerbaarheid voor het onderzoek gekozen om slechts één sessie (les) voor het spelen met Metriek in te ruimen. De beperkte bandbreedte van de drie toets-scores is mogelijk aan voornoemde oorzaken te wijten. In combinatie met de gemeten in-groep varianties verhindert of bemoeilijkt deze beperkte bandbreedte het opsporen van mogelijke statistisch significante verschillen (= ruis in de data). Het voornoemde wordt mogelijkwerwijs nog versterkt door de derde beperking in dit interventieonderzoek, namelijk beperkte betrouwbaarheid van de – zelf ontwikkelde – kennistoetsen zoals bleek uit de betrouwbaarheidsanalyses.

De twee laatste beperkingen vloeien voort uit de beperkte controleerbaarheid van ecologisch valide onderzoek en de gebruikte game-ontwikkelomgeving. Een geheel onverwachte bevinding dat de KCR-groep een significant hogere retentietoetsscore heeft dan posttoetsscore. Het kan zijn dat de leerlingen uit de KCR-groep ná de posttoets – onbedoeld – extra informatie hebben gekregen, waardoor ze beter scoorden op de retentietoets. Het is mogelijk dat leerlingen door zijn gegaan met het hoofdstuk over het metriek stelsel en het berekenen van de omtrek. Dit kan verklaren waarom dat leerlingen uit de KCR-groep hoger scoorden op de retentietoets dan op de posttoets. Dit kan bovendien verklaren waarom de KCR-groep op de retentietoets significant hoger scoorde dan de EF-groep. Tot slot lijkt het erop dat de leermaterialen (feedback, self-explanation en theorie) anders zijn gebruikt dan werd beoogd. Zo kunnen feedback, self-explanation en theorie niet goed zijn gelezen, gebruikt of zelfs zijn overgeslagen. Helaas zijn binnen Seppo geen mogelijkheden om het feitelijke gebruik van de leermaterialen uit de logging te destilleren. Eerder onderzoek laat bijvoorbeeld zien dat leerlingen hun studeergedrag veranderen als ze feedback verwachten (Maier et al., 2016) of dat ze de feedback niet lezen (Bokhove & Drijvers, 2012). Binnen het onderzoek is Seppo gebruikt met een snelle inleercurve opdat de onderzoeker zelfstandig in staat was om de game-varianten te ontwikkelen voor het

interventie-onderzoek. Hierdoor moesten concessies worden gedaan ten aanzien van de te verzamelen in-game data.

5.3 Aanbevelingen voor vervolgonderzoek

Allereerst is in een vervolgonderzoek van belang dat de onderzoeker beter weet wat de voorkennis van de leerlingen is. Dit pleit er voor om alvorens een onderzoek te starten, eerst te controleren of leerlingen wellicht toch niet al over voorkennis beschikken op dit onderwerp.

Ten tweede moet goed worden nagedacht over de omvang van de serious game, zodat alle leerstof goed kan worden ingeoeft, hierbij kan gedacht worden aan een serious game met een langere aaneengesloten speelduur. Door een langere aaneengesloten speelduur kan worden gezorgd voor genoeg oefenmateriaal en voldoende controleerbaarheid van de onderzoeksetting.

Ten derde is het gebruik van gevalideerde kennistoetsen van belang om de effectiviteit van de serious game goed te kunnen meten.

Ten vierde is extra afstemming tussen alle betrokkenen belangrijk na de afname van de posttoets, zodat leerlingen niet zelfstandig verder kunnen werken met het onderwerp. Dit kan bijvoorbeeld door meteen door te gaan met een ander onderwerp en restricties te zetten op het hetgeen de leerlingen online kunnen oefenen. Een andere optie is om leerlingen duidelijk te maken dat de stof van de serious game later in het jaar niet terugkomt, zodat voor de leerlingen geen prikkel is om zelfstandig aan de slag te gaan met de stof.

Tot slot kan het aangaan van een samenwerking met een bedrijf dat gespecialiseerd is in het ontwikkelen van serious games voor het vak wiskunde het mogelijk maken om specifiekere in-game data te verzamelen die behulpzaam zijn bij de interpretatie van de performance-gegevens van de leerlingen op het game-onderwerp (verbeterde logging). Zo kan bijvoorbeeld gedacht worden aan het inspreken van de theorieblokken (bijvoorbeeld als kennisclip) en de feedback. Leerlingen moeten dan verplicht naar de theorie en feedback luisteren, voordat ze verder kunnen in de game. Op deze manier lijkt de kans groter dat leerlingen echt aandacht besteden aan de feedback en aan de theorieblokken.

Om de effectieve leerstrategie self-explanation in combinatie met een uitgewerkt voorbeeld beter aan bod te laten, moeten in de kennisclips uitgewerkte voorbeelden van een realistisch probleem worden gegeven. Het blijkt dat het combineren van een realistisch probleem en een uitgewerkt voorbeeld samen voor betere leerresultaten kan zorgen (van Merriënboer & Kirschner, 2017). Het visueel maken van een goed uitgewerkt voorbeeld kan helpen bij het aanleren van (complexe) vaardigheden en geven de leerlingen een handvat om de opdracht aan zichzelf uit te leggen (van Merriënboer & Kirschner, 2017). Het werken met uitgewerkte voorbeelden in combinatie met self-explanation is nog effectiever voor leren (Barton, 2019; Atkinson et al., 2000). Dit vergt meer middelen en bijdragen vanuit verschillende disciplines expertise dan bij dit scriptietraject beschikbaar waren. Zo zouden een geavanceerder game-ontwikkelplatform, intensievere samenwerking tussen

verschillende disciplines (game design, instructional design, vakdidactiek, storytelling) zowel ten gunste komen aan de educatieve kwaliteit van de game alsook aan de kwaliteit van het onderzoek.

5.4 Conclusie

Op basis van dit onderzoek wordt geconcludeerd dat MeTriek de potentie heeft om leerlingen te laten leren. Feedbacktype KR en EF zorgen op korte termijn voor hogere leerresultaten en KR en KCR zorgen op langere termijn voor hogere leerresultaten. De drie feedbacktypes blijken onderling niet van elkaar te verschillen op hun leereffecten (op een bepaald moment en door de tijd). Dit resultaat komt niet overeen met eerder onderzoek. Beter controleerbaar vervolgonderzoek waarbij ondermeer beter gekeken wordt naar de voorkennis, gebruik van in-game data van de meer geavanceerd ontwikkelde game met bijvoorbeeld gesproken feedback, en gevalideerde toetsen kan onderzoek naar feedbacktypes in serious games verbeteren. Deze aanpassingen zullen helpen een antwoord te vinden op de vraag of elaboratieve feedback in wiskunde (serious) games zorgen voor een beter leereffect dan KCR en KR-feedback.

Referenties

- Anderson, J., Fincham, J., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of Experimental Psychology-Learning Memory and Cognition*, 23(4), 932-945. doi:10.1037//0278-7393.23.4.932
- Annetta, L. A. (2010). The "I's" have it: A framework for serious educational game design. *Review of General Psychology*, 14(2), 105-113. doi:10.1037/a0018985
- Atkinson, R. K., & Renkl, A. (2007). Interactive example-based learning environments: Using interactive elements to encourage effective processing of worked examples. *Educational Psychology Review*, 19(3), 375-386. doi:10.1007/s10648-007-9055-2
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of Educational Research*, 70(2), 181-214. doi:10.3102/00346543070002181
- Barton, C. (2019). *Volgens Barton Deel 1: Lesgeven in wiskunde aan de hand van wetenschap, experts, en 12 jaar aan mislukkingen*. Culemborg: Phronese.
- Bokhove, C., & Drijvers, P. (2012). Effects of feedback in an online algebra intervention. *Technology, Knowledge and Learning*, 17(1-2), 43-59. doi:10.1007/s10758-012-9191-8
- Bruyckere, P. de, (2017) Een wat langer stuk over de taxonomie van Bloom, <https://pedrodebruyckere.blog/2017/04/17/een-wat-langer-stuk-over-de-taxonomie-van-bloom/>
- Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology*, 105(2), 290-298. doi:10.1037/a0031026
- Clark, D. B., Tanner-Smith, E. E., & Killingsworth, S. S. (2016). Digital games, design, and learning: A systematic review and meta-analysis. *Review of educational research*, 86(1), 79-122. doi:10.3102/0034654315582065
- Connolly, T. M., Boyle, E. A., MacArthur, E., Hainey, T., & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & education*, 59(2), 661-686. doi:10.1016/j.compedu.2012.03.004
- Cresswell, J. W. (2014). *Educational research: planning, conducting, and evaluating quantitative and qualitative research* (4e druk). Harlow, Verenigd Koninkrijk: Pearson Education.
- Dalziel, J. (2001). Enhancing web-based learning with computer assisted assessment: Pedagogical and technical considerations.
- Dempsey, J. V., Driscoll, M. P., & Lichtfield, B. C. (1993). Feedback, retention, discrimination error, and feedback study time. *Journal of Research on Computing in Education*, 25 (3), 303-326. doi:10.1080/08886504.1993.10782053

- Drijvers, P. (2013). Digital technology in mathematics education: Why it works (or doesn't) [la tecnología digital en educación matemática: Por qué funciona (o no)]. *Pna*, 8(1), 1-20.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58.
doi:10.1177/1529100612453266
- Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175-191. doi:10.3758/BF03193146
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics* (4th ed.). London: SAGE Publications Ltd.
- Finneran, C. M., & Zhang, P. (2005). Flow in computer-mediated environments: promises and challenges. *Communications of the association for information systems*, 15(1), 4.
doi:10.17705/1CAIS.01504
- Fyfe, E. R. (2016). Providing feedback on computer-based algebra homework in middle-school classrooms. *Computers in Human Behavior*, 63, 568-574. doi:10.1016/j.chb.2016.05.082
- Fyfe, E. R., Rittle-Johnson, B., & DeCaro, M. S. (2012). The effects of feedback during exploratory mathematics problem solving: Prior knowledge matters. *Journal of Educational Psychology*, 104(4), 1094-1108. doi:10.1037/a0028389
- Gaggi, O., & Petenazzi, G. (2019). A digital platform for teaching mathematics. Paper presented at the 37-42. doi:10.1145/3342428.3342666
- Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE)*, 1(1), 20-20. doi:10.1145/950566.950595
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1), 81-112. doi:10.3102/003465430298487
- Hattie, J. A., & Marsh, H. W. (1995). Future research in self-concept. *Handbook on self-concept*, 421-463.
- Hsieh, I. L. G., & O'Neil Jr, H. F. (2002). Types of feedback in a computer-based collaborative problem-solving group task. *Computers in Human Behavior*, 18(6), 699-715.
doi:10.1016/S0747-5632(02)00025-0
- Hsu, C. Y., & Tsai, C. C. (2013). Examining the effects of combining self-explanation principles with an educational game on learning science concepts. *Interactive Learning Environments*, 21(2), 104-115. doi:10.1080/10494820.2012.705850
- Johnson, C. I., Bailey, S. K., & Van Buskirk, W. L. (2017). Designing effective feedback messages in serious games and simulations: a research review. In *Instructional Techniques to Facilitate Learning and Motivation of Serious Games* (pp. 119-140). Springer, Cham.

- Kebritchi, M., Hirumi, A., & Bai, H. (2010). The effects of modern mathematics computer games on mathematics achievement and class motivation. *Computers & education*, 55(2), 427-443. doi:10.1016/j.compedu.2010.02.007
- Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and higher education*, 8(1), 13-24. doi:10.1016/j.iheduc.2004.12.001
- Kiili, K. J. M., Devlin, K., & Multisilta, J. (2015). Editorial: Is game-based math learning finally coming of age? *International Journal of Serious Games*, 2(4), 1-4. doi:10.17083/ijsg.v2i4.109
- Kiili, K., Moeller, K., & Ninaus, M. (2018). Evaluating the effectiveness of a game-based rational number training - in-game metrics as learning indicators. *Computers and Education*, 120, 13-28. doi:10.1016/j.compedu.2018.01.012
- Kouwenhoven, M. (2019). *De effectiviteit van Verschillende Feedbackvormen op de Leerresultaten van een Serious Game in het Basisonderwijs*. Open Universiteit.
- Mason, B. J., & Bruning, R.H. (2001). *Providing feedback in computer-based instruction: What the research tells us*. CLASS Research Report No. 9. Center for Instructional Innovation, University of Nebraska-Lincoln.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901-918. doi:10.1037/a0037123
- Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers and Education*, 95(95), 85-98. doi.org/10.1016/j.compedu.2015.12.002
- Mayer, R. E. (2014). *Computer games for learning: An evidence-based approach*. MIT Press.
- Ministerie van Onderwijs, Cultuur en Wetenschap. (2019, 12 16). *Info over wiskundevaardigheden*. Opgehaald van Website van het Ministerie van Onderwijs, Cultuur en Wetenschap: <https://www.ocwincijfers.nl/verantwoord-begroten/internationaal/eu2020-strategie/percentage-15-jarige-leerlingen-met-lage-wiskunde-vaardigheden>
- Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multi-media learning. In H. M. Niegemann, D. Leutner, & R. Brünken (Eds.), *Instructional Design for Multimedia learning* (pp. 181-195). Münster, New York: Waxman.
- Parsons, S., & Bynner, J. (2005). Does numeracy matter more?.
- Phye, G. D. (1979). The processing of informative feedback about multiple-choice test performance. *Contemporary Educational Psychology*, 4 (1979), pp. 381-394. doi:10.1016/0361-476X(79)90057-2
- Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153-189. doi:10.3102/0034654307313795

- Sweller, J., van Merriënboer, Jeroen J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 Years later. *Educational Psychology Review*, 31(2), 261-292.
doi:10.1007/s10648-019-09465-5
- ter Vrugte, J., & de Jong, T. (2017). Self-explanations in game-based learning: From tacit to transferable knowledge. In *Instructional techniques to facilitate learning and motivation of serious games* (pp. 141-159). Springer, Cham.
- Timmers, C. F., & Veldkamp, B. P. (2011). Attention paid to feedback provided by a computer-based assessment for learning on information literacy. *Computers & Education*, 56, 923-930.
doi:10.1016/j.compedu.2010.11.007
- van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of educational research*, 85(4), 475-511. doi:10.3102/0034654314564881
- van Merriënboer, J. J., & Kirschner, P. A. (2017). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Routledge.
- van Rooij, A., Dalinghaus, N., & van den Eijnden, R. (2018). *Info over (on)gezond gamegedrag*. Opgehaald van Website van het Trimbos-instituut): <https://www.trimbos.nl/docs/76e0b500-bc80-41bc-ba86-2599015e36d2.pdf>
- Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.
- Wilson, A. J., Dehaene, S., Dubois, O., & Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind, Brain and Education*, 3(4), 224-234. doi:10.1111/j.1751-228X.2009.01075.x
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of educational psychology*, 105(2), 249-265. doi:10.1037/a0031311
- Wouters, P., & van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education*, 60(1), 412-425.
doi:10.1016/j.compedu.2012.07.018

Bijlage

Bijlage 1. Vragenlijst kwaliteitsmeting serious game (Kouwenhoven, 2019)

Identity

1. Ik had het gevoel onderdeel te zijn van de game
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik vind dat ik een belangrijke taak vervulde binnen de game
[1: Geheel mee oneens – 5: Geheel mee eens]

Immersion

1. Ik heb erg genoten van het doen van deze game
[1: Geheel mee oneens – 7: Geheel mee eens]
 2. Deze game was leuk om te doen
[1: Geheel mee oneens – 7: Geheel mee eens]
 3. Ik denk dat het doen van deze game saai was*
[1: Geheel mee oneens – 7: Geheel mee eens]
 4. Deze game heeft mijn aandacht helemaal niet getrokken*
[1: Geheel mee oneens – 7: Geheel mee eens]
 5. Ik zou deze game als zeer interessant willen omschrijven
[1: Geheel mee oneens – 7: Geheel mee eens]
 6. Toen ik deze game deed, dacht ik hoe plezierig het was
[1: Geheel mee oneens – 7: Geheel mee eens]
 7. Ik heb veel inspanning geleverd voor deze game
[1: Geheel mee oneens – 7: Geheel mee eens]
 8. Ik heb niet hard mijn best gedaan om deze game goed te doen*
[1: Geheel mee oneens – 7: Geheel mee eens]
 9. Ik heb heel hard mijn best gedaan in deze game
[1: Geheel mee oneens – 7: Geheel mee eens]
 10. Het was belangrijk voor me om de game goed te doen
[1: Geheel mee oneens – 7: Geheel mee eens]
 11. Ik heb weinig energie in de game gestoken*
[1: Geheel mee oneens – 7: Geheel mee eens]
-
1. Ik voelde me prettig uitgedaagd
[1: Geheel niet waar – 7: Helemaal waar]
 2. Mijn gedachten verliepen soepel
[1: Geheel niet waar – 7: Helemaal waar]
 3. Mijn hoofd was helemaal helder
[1: Geheel niet waar – 7: Helemaal waar]
 4. Ik was helemaal verdiept in wat ik deed
[1: Geheel niet waar – 7: Helemaal waar]
 5. De juiste gedachten kwamen vanzelf
[1: Geheel niet waar – 7: Helemaal waar]
 6. Bij iedere stap wist ik wat ik moest doen
[1: Geheel niet waar – 7: Helemaal waar]
 7. Ik had het gevoel dat ik alles onder controle had
[1: Geheel niet waar – 7: Helemaal waar]
 8. Ik was in gedachten verzonken
[1: Geheel niet waar – 7: Helemaal waar]
 9. Ik had er geen idee van hoe snel de tijd liep
[1: Geheel niet waar – 7: Helemaal waar]
 10. Ik vond het gemakkelijk om me te concentreren

[1: Geheel niet waar – 7: Helemaal waar]

1. Ik vind dat ik voldoende controle heb binnen de game
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik vind de uitleg bij de besturing van de game duidelijk
[1: Geheel mee oneens – 5: Geheel mee eens]
3. Het is voldoende duidelijk wat je in de game moest doen
[1: Geheel mee oneens – 5: Geheel mee eens]
4. Ik vind de uitleg bij de opdrachten in de game duidelijk
[1: Geheel mee oneens – 5: Geheel mee eens]
5. Ik vind dat de game te lang duurt*
[1: Geheel mee oneens – 5: Geheel mee eens]

Interactivity

1. In deze game is voldoende mogelijkheid om vragen te stellen of opmerkingen te maken
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik heb het contact dat ik tijdens de game had als prettig ervaren
[1: Geheel mee oneens – 5: Geheel mee eens]
3. Ik zou willen dat ik meer contact kon hebben tijdens het spel*
[1: Geheel mee oneens – 5: Geheel mee eens]

Increasing complexity

1. Hoe mentaal belastend waren de opdrachten in de game?*
 2. Hoe fysiek belastend waren de opdrachten in de game?*
 3. Hoe gehaast was het tempo van de opdrachten in de game?*
 4. Hoe succesvol was je in het doen van de opdrachten in de game?
 5. Hoe hard moest je je inspannen om de opdrachten in de game succesvol te kunnen doen?*
 6. Hoeveel negatieve gevoelens had je tijdens de opdrachten in de game?*
-
1. Ik vind de game gemakkelijk
[1: Geheel mee oneens – 5: Geheel mee eens]
 2. Ik vind dat de opdrachten in de game in een goede volgorde staan
[1: Geheel mee oneens – 5: Geheel mee eens]

Informed teaching

1. In de game zouden meer terugkoppeling en aanwijzingen gegeven moeten worden*
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik vind de terugkoppelingen en aanwijzingen in de game goed bruikbaar
[1: Geheel mee oneens – 5: Geheel mee eens]
3. Ik vind dat de ingebouwde begeleiding bij de opdrachten in de game goed is
[1: Geheel mee oneens – 5: Geheel mee eens]
4. Ik wist waarop ik beoordeeld zou worden bij de game
[1: Geheel mee oneens – 5: Geheel mee eens]

Instructional

1. In de game zou meer kennis over lengtematen en omtrek verwerkt moeten zijn*
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Voor het uitvoeren van de opdrachten kreeg ik voldoende informatie
[1: Geheel mee oneens – 5: Geheel mee eens]
3. In deze game wordt de kennis over lengtematen en omtrek voldoende getoetst
[1: Geheel mee oneens – 5: Geheel mee eens]

*Voor deze vragen is de score omgekeerd

Bijlage 2. Aangepaste vragenlijst kwaliteitsmeting serious game

Identity

1. Ik had het gevoel onderdeel te zijn van de game
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik vind dat ik een belangrijke taak vervulde binnen de game
[1: Geheel mee oneens – 5: Geheel mee eens]

Immersion

1. Ik heb erg genoten van het doen van deze game
[1: Geheel mee oneens – 7: Geheel mee eens]
2. Deze game was leuk om te doen
[1: Geheel mee oneens – 7: Geheel mee eens]
3. Ik vind de game saai om te doen*
[1: Geheel mee oneens – 7: Geheel mee eens]
4. Deze game heeft mijn aandacht niet getrokken*
[1: Geheel mee oneens – 7: Geheel mee eens]
5. Ik vind deze game zeer interessant
[1: Geheel mee oneens – 7: Geheel mee eens]
6. Tijdens het spelen van de game, dacht ik hoe plezierig het was
[1: Geheel mee oneens – 7: Geheel mee eens]
7. Ik heb veel inspanning geleverd voor deze game
[1: Geheel mee oneens – 7: Geheel mee eens]
8. Ik heb niet hard mijn best gedaan om deze game goed te doen*
[1: Geheel mee oneens – 7: Geheel mee eens]
9. Ik heb heel hard mijn best gedaan in deze game
[1: Geheel mee oneens – 7: Geheel mee eens]
10. Het was belangrijk voor me om de game goed te doen
[1: Geheel mee oneens – 7: Geheel mee eens]
11. Ik heb weinig energie in de game gestoken*
[1: Geheel mee oneens – 7: Geheel mee eens]

Flow

1. Ik voelde me prettig uitgedaagd
[1: Geheel niet waar – 7: Helemaal waar]
2. Mijn gedachten verliepen soepel
[1: Geheel niet waar – 7: Helemaal waar]
3. Mijn hoofd was helemaal helder
[1: Geheel niet waar – 7: Helemaal waar]
4. Ik was helemaal verdiept in wat ik deed

[1: Geheel niet waar – 7: Helemaal waar]

5. De juiste gedachten kwamen vanzelf

[1: Geheel niet waar – 7: Helemaal waar]

6. Bij iedere stap wist ik wat ik moest doen

[1: Geheel niet waar – 7: Helemaal waar]

7. Ik had het gevoel dat ik alles onder controle had

[1: Geheel niet waar – 7: Helemaal waar]

8. Ik was tijdens het spelen van de game met mijn gedachten bij de game.

[1: Geheel niet waar – 7: Helemaal waar]

9. Ik had er geen idee van hoe snel de tijd ging.

[1: Geheel niet waar – 7: Helemaal waar]

10. Ik vond het gemakkelijk om me te concentreren

[1: Geheel niet waar – 7: Helemaal waar]

Studeerbaarheidsvragenlijst

1. Ik vind dat ik voldoende controle heb binnen de game

[1: Geheel mee oneens – 5: Geheel mee eens]

2. Ik vind de uitleg bij de besturing van de game duidelijk

[1: Geheel mee oneens – 5: Geheel mee eens]

3. Het is voldoende duidelijk wat je in de game moest doen

[1: Geheel mee oneens – 5: Geheel mee eens]

4. Ik vind de uitleg bij de opdrachten in de game duidelijk

[1: Geheel mee oneens – 5: Geheel mee eens]

5. Ik vind dat de game te lang duurt*

[1: Geheel mee oneens – 5: Geheel mee eens]

Interactivity

1. In deze game is voldoende mogelijkheid om vragen te stellen of opmerkingen te maken

[1: Geheel mee oneens – 5: Geheel mee eens]

2. Ik heb het contact dat ik tijdens de game had als prettig ervaren

[1: Geheel mee oneens – 5: Geheel mee eens]

3. Ik zou willen dat ik meer contact kon hebben tijdens het spel*

[1: Geheel mee oneens – 5: Geheel mee eens]

Increasing complexity

1. Hoeveel energie heeft het spelen van de game je gekost?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]
2. Hoe fysiek belastend waren de opdrachten in de game?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]
3. Hoe gehaast was het tempo van de opdrachten in de game?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]
4. Hoe succesvol was je in het doen van de opdrachten in de game?
[1: Zeer laag belastend – 7: Zeer hoog belastend]
5. Hoe hard moest je je inspannen om de opdrachten in de game succesvol te kunnen doen?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]
6. Hoeveel negatieve gevoelens had je tijdens de opdrachten in de game?*[
[1: Zeer laag belastend – 7: Zeer hoog belastend]

Studeerbaarheidslijst

7. Ik vind de game gemakkelijk
[1: Geheel mee oneens – 5: Geheel mee eens]
8. Ik vind dat de opdrachten in de game in een goede volgorde staan
[1: Geheel mee oneens – 5: Geheel mee eens]

Informed teaching

1. In de game zouden meer terugkoppelingen (feedback) en aanwijzingen gegeven moeten worden*[
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Ik vind de terugkoppelingen (feedback) en aanwijzingen in de game goed bruikbaar
[1: Geheel mee oneens – 5: Geheel mee eens]
3. Ik vind dat de ingebouwde begeleiding bij de opdrachten in de game goed is
[1: Geheel mee oneens – 5: Geheel mee eens]
4. Ik wist waarop ik beoordeeld zou worden bij de game
[1: Geheel mee oneens – 5: Geheel mee eens]

Instructional

1. In de game zou meer kennis over lengtematen en omtrek verwerkt moeten zijn*[
[1: Geheel mee oneens – 5: Geheel mee eens]
2. Voor het uitvoeren van de opdrachten kreeg ik voldoende informatie
[1: Geheel mee oneens – 5: Geheel mee eens]
3. In deze game wordt de kennis over lengtematen en omtrek voldoende getoetst
[1: Geheel mee oneens – 5: Geheel mee eens]

* voor deze vragen is de score omgekeerd

Bijlage 3 Verantwoording serious game volgens het framework van Annetta (2010)

Identity

In een kwalitatief goed educatief spel voelt elke speler zich een uniek personage in de game. In de meeste games wordt dit georganiseerd door een Avatar te gebruiken. Een Avatar is een gepersonaliseerde representatie of een karakter die de speler van de serious game vertegenwoordigt in-game. In Seppo is het niet mogelijk om de leerling een Avatar te laten kiezen. Wel is het mogelijk om een verhaallijn aan het spel toe te voegen. Identity wordt in MeTriek gecreëerd door de leerling bij de start van het spel mee te nemen in het verhaal. WisKid, de NPC in het spel vraagt de leerlingen om hem te helpen met het oplossen van wiskundige vragen in en rondom de school. In de game worden ook afbeeldingen gebruikt uit de school van de leerlingen. Kiili (2005) beschrijft ook dat een verhaallijn helpt om leerlingen het gevoel van identity en immersion te geven. Door afbeeldingen van de onderzoeksschool te gebruiken heeft de onderzoeker geprobeerd om de identity van leerlingen in de game te verhogen.

Immersion

Immersion betekent dat een speler betrokken is met de inhoud van het spel en zijn of haar aanwezigheid voelt in een bepaalde context. Spelers worden dan betrokken en intrinsiek gemotiveerd om de game succesvol af te ronden. Dit gevoel van immersion kan worden gecreëerd door een verhaallijn in te bouwen waar leerlingen bij betrokken zijn (Kiili, 2005).

Als een speler betrokken en gemotiveerd is kan de speler in een staat van flow terecht komen. De leerling raakt in MeTriek betrokken doordat de omgeving in de game plaatsvindt op bekend terrein (de school). Volgens Kiili (2005) is het geven van feedback en het duidelijk stellen van doelen ook essentieel voor flow. De leerlingen krijgen in MeTriek meteen feedback na het geven van een antwoord en het doel van het spel wordt tijdens het startverhaal en tijdens de vragen uitgelegd door WisKid.

Interactivity

Seppo heeft de mogelijkheid om chatboxen te gebruiken. Eén chatbox voor contact met instructeurs is in de game aanwezig. De leerlingen kunnen een vraag stellen aan WisKid via deze chatbox. Ze krijgen echter tijdens het spelen geen antwoord, waardoor dat de resultaten niet beïnvloed kunnen worden. Seppo heeft voor de rest weinig mogelijkheden om de interactivity te verhogen via een desktop of laptop. Nieuwe ontwikkelingen hebben het mogelijk gemaakt om met de Seppo App naar buiten te gaan en de game buiten te spelen. Deze optie was (helaas) nog niet beschikbaar en mogelijk toen dit onderzoek werd gestart.

Increasing complexity

Het is in Seppo mogelijk om meerdere levels toe te voegen. Een goede game bestaat uit meerdere levels die oplopen in moeilijkheidsgraad. In de game MeTriek lopen de opdrachten binnen de vier levels ook op in moeilijkheidsgraad. Dit gebeurt door de complexere rekenstrategieën in de hogere levels aan te bieden. Leerlingen kunnen pas door naar het volgende level als alle opdrachten zijn gemaakt. Deze opdrachten hoeven niet goed gemaakt te zijn. In het eerste level krijgen leerlingen instructie over het verschil tussen eenheden en grootheden, in het tweede level leren leerlingen de stappen om te rekenen met lengtematen, in het derde level leren leerlingen de omtrek te berekenen. De leerlingen sluit de game af met een combinatie van alle opdrachten, waardoor dat de complexiteit ook hoger is.

Informed teaching

Door het gebruik van Seppo heeft de onderzoeker de mogelijkheid om verschillende data te verzamelen, zoals bijvoorbeeld de speeltijd, de scores en een overzicht van de gegeven antwoorden. Het is ook mogelijk om de chatgeschiedenis tussen de instructeur (WisKid) en de leerlingen op te vragen

Instructional

De serious game MeTriek is ter vervanging een deel uit de wiskundemethode Getal en Ruimte 12^e editie. De leerlingen krijgen in het spel over verschillende onderwerpen een instructie en maken verschillende vragen. De leerlingen krijgen in het begin van elk level nieuwe instructie van WisKid en tijdens de opdrachten legt WisKid ook uit wat de leerlingen moeten doen. De instructie wordt met behulp van tekst en afbeeldingen gegeven.

Bijlage 4 kennistoetsen

MeTriek

versie A

Opdracht 1

Welke waarde hoort er bij 'hecto'?

- A. 0,1 B. 10 C. 100 D. 1000

Opdracht 2

Welke eenheid hoort er op de plaats van de te staan?

Snijdt 100 boter in blokjes.

- A. mL
B. minuten
C. g
D. °C

Opdracht 3

De *liter* is een eenheid van

Antwoord:

Opdracht 4

Zet de eenheden op volgorde van groot naar klein

dm	hm	cm	km	dam	mm	m
----	----	----	----	-----	----	---

Antwoord:

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

Opdracht 5

Bereken en vul in:

25 m = mm

Opdracht 6

Bereken en vul in:

10000 m = dam

Opdracht 7

Bereken en vul in:

15,2 mm = cm

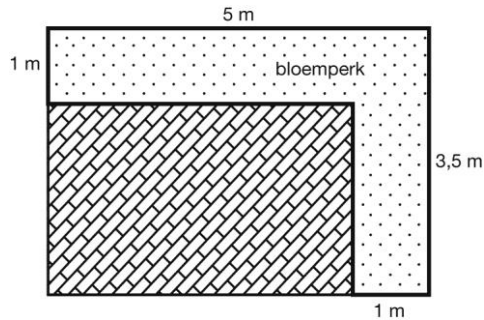
Opdracht 8

Bereken en vul in:

6,2 km = m

Opdracht 9

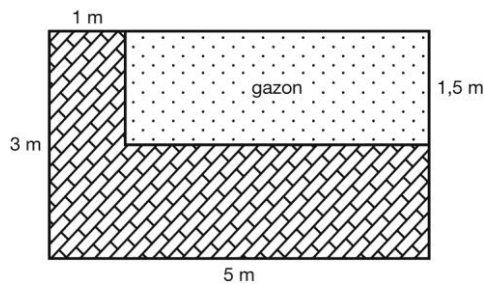
Bereken de omtrek van de totale rechthoek.



Antwoord:.....

Opdracht 10

Bereken de omtrek van het gazon.



Antwoord:

Opdracht 11

Je hebt een stuk land van 17 meter bij 18 meter.
Wat is de omtrek van het stuk land?

Opdracht 12

Een stuk land is 12 meter bij 700 cm.
Wat is de omtrek van het stuk land in cm?

MeTriek

versie B

Opdracht 1

Welke waarde hoort er bij 'deca'?

- A. 0,1 B. 10 C. 100 D. 1000

Opdracht 2

Welke eenheid hoort er op de plaats van de te staan?

Voeg nu 50 water toe.

- A. mL
B. minuten
C. g
D. °C

Opdracht 3

De *Kilo* is een eenheid van

Antwoord:

Opdracht 4

Zet de eenheden op volgorde van groot naar klein

cm	hm	dm	km	dam	mm	m
----	----	----	----	-----	----	---

Antwoord:

--	--	--	--	--	--	--

Opdracht 5

Bereken en vul in:

4,3 m = mm

Opdracht 6

Bereken en vul in:

750 dm = km

Opdracht 7

Bereken en vul in:

25,3 mm = cm

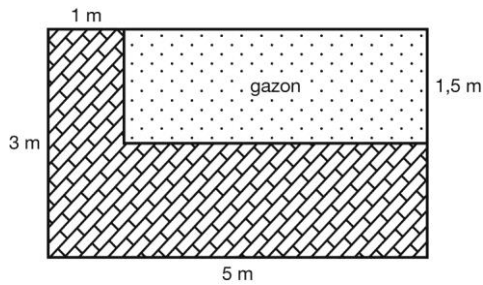
Opdracht 8

Bereken en vul in:

3,4 km = m

Opdracht 9

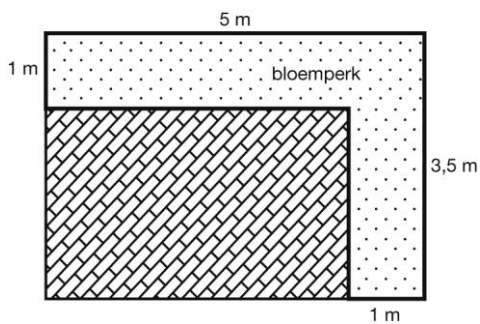
Bereken de omtrek van de totale rechthoek.



Antwoord:.....

Opdracht 10

Bereken de omtrek van het bloemperk.



Antwoord:

Opdracht 11

Je hebt een stuk land van 14 meter bij 13 meter.

Wat is de omtrek van het stuk land?

Opdracht 12

Een stuk land is 17 meter bij 900 cm.

Wat is de omtrek van het stuk land in cm?

MeTriek

versie C

Opdracht 1

Welke waarde hoort er bij 'kilo'?

- A. 0,1 B. 10 C. 100 D. 1000

Opdracht 2

Welke eenheid hoort er op de plaats van de te staan?

Bak ze in de oven in 20 – 25

- A. mL
B. minuten
C. g
D. °C

Opdracht 3

De hectometer is een eenheid van

Antwoord:

Opdracht 4

Zet de eenheden op volgorde van groot naar klein

mm	hm	km	dm	dam	cm	m
----	----	----	----	-----	----	---

Antwoord:

<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
----------------------	----------------------	----------------------	----------------------	----------------------	----------------------	----------------------

Opdracht 5

Bereken en vul in:

82 m = mm

Opdracht 6

Bereken en vul in:

25000 m = dam

Opdracht 7

Bereken en vul in:

33,6 mm = cm

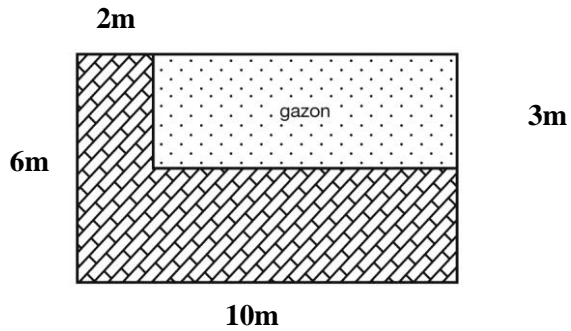
Opdracht 8

Bereken en vul in:

5,8 km = m

Opdracht 9

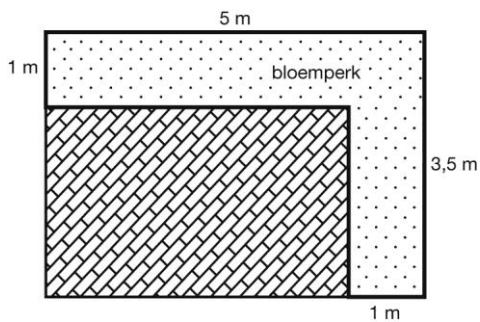
Bereken de omtrek van de totale rechthoek.



Antwoord:.....

Opdracht 10

Bereken de omtrek van het bloemperk.



Antwoord:

Opdracht 11

Je hebt een stuk land van 24 meter bij 36 meter.

Wat is de omtrek van het stuk land?

Opdracht 12

Een stuk land is 21 meter bij 400 cm.

Wat is de omtrek van het stuk land in cm?